

# American Political Science Review

Volume 99 | Number 3 | August 2005

CUK-HO5430-13-KP3868



## INSIDE THIS ISSUE:

### NOTES FROM THE EDITOR

#### ARTICLES

- Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina. **Susan C. Stokes** ..... 315
- A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective. **Tulia G. Falletti** ..... 327
- Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success. **Bonnie M. Meguid** ..... 347
- Uncovering Evidence of Conditional Party Government: Reassessing Majority Party Influence in Congress and State Legislatures. **William T. Bianco and Itai Sened** ..... 367
- Racism Revised: Courts, Labor Law, and the Institutional Construction of Racial Animus. **Paul Frymer** ..... 377
- Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure. **B. Peter Rosendorff** ..... 389
- Reading Habermas in Anarchy: Multilateral Diplomacy and Global Public Spheres. **Jennifer Mitzen** ..... 407
- Prenegotiation Public Commitment in Domestic and International Bargaining. **Bahar Leventoglu and Ahmer Tarar** ..... 419
- Nested Analysis as a Mixed-Method Strategy for Comparative Research. **Evan S. Lieberman** ..... 431

#### FORUM

- No Rest for the Democratic Peace. **David Kinsella** ..... 451
- Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace. **Branislav L. Slantchev, Anna Alexandrova, and Erik Gartzke** ..... 457
- Three Pillars of the Liberal Peace. **Michael W. Doyle** ..... 461
- Explaining the Democratic Peace **Sebastian Rosato** ..... 467



## EDITOR

Lee Sigelman  
*Department of Political Science  
The George Washington University  
Washington, DC 20052*

## EXECUTIVE COMMITTEE

### American Politics

Robert Stein  
*Rice University*

### Comparative Politics

Thomas Remington  
*Emory University*

### International Relations

Edward Mansfield  
*University of Pennsylvania*

### Political Theory

Ruth W. Grant  
*Duke University*

### At-Large

K.C. Morrison  
*University of Missouri*  
M. Elizabeth Sanders  
*Cornell University*

## EDITORIAL BOARD

Martha Ackelsberg

*Smith College*

James Adams

*University of California,*

*Santa Barbara*

Kathleen Bawn

*University of California,*

*Los Angeles*

Janet Box-Steffensmeier

*The Ohio State University*

Cathy Cohen

*University of Chicago*

Timothy E. Cook

*Louisiana State University*

Neta Crawford

*Brown University*

Ivor Crewe

*University of Essex*

Darren Davis

*Michigan State University*

Jon Elster

*Columbia University*

Diana Evans

*Trinity College*

Steven Finkel

*University of Virginia*

Kim L. Fridkin

*Arizona State University*

Robert Goodin

*Australian National*

*University*

Peter A. Hall

*Harvard University*

Richard Herrmann

*The Ohio State University*

Rodney Hero

*University of Notre Dame*

James Johnson

*University of Rochester*

Jane Junn

*Rutgers University*

George Klosko

*University of Virginia*

Arthur Lupia

*University of Michigan*

Forrest Maltzman

*The George Washington*

*University*

Kenneth Mayer

*University of Wisconsin*

Kirstie McClure

*University of California,*

*Los Angeles*

Kenneth Meier

*Texas A&M University*

*and Cardiff University*

James D. Morrow

*University of Michigan*

Rebecca Morton

*New York University*

Robert Price

*University of California,*

*Berkeley*

Kathryn Sikkink

*University of Minnesota*

Beth Simmons

*Harvard University*

Jack Snyder

*Columbia University*

Janice Gross Stein

*University of Toronto*

Sven Steinmo

*University of Colorado*

Shannon Stimson

*University of California,*

*Berkeley*

Leslie Paul Thiele

*University of Florida*

Daniel Treisman

*University of California,*

*Los Angeles*

Clyde Wilcox

*Georgetown University*

John Wright

*The Ohio State University*

Bernard Yack

*Brandeis University*

Langche Zeng

*The George Washington*  
*University*

## EDITORIAL STAFF

Elizabeth Cook  
*Assistant Editor*

Jennifer Deets  
Elizabeth Franker  
Lee Michael  
Jennie Schulze  
*Editorial Assistants*

**Association Office Address and Membership Information:** American Political Science Association individual membership dues are: Regular Members with income \$100,000+, \$190; \$80,000–\$99,999, \$163; \$60,000–\$79,999, \$146; \$50,000–\$59,999, \$136; \$40,000–\$49,999, \$115; Under \$40,000, \$78; Student Members \$37; Retired over \$25,000, \$52; Retired under \$25,000, \$32; Life Members, \$3,000.00; Unemployed members, \$37.00; Associate Member with one journal selection, \$48. The amount of the dues allocated for a subscription is \$5 for an individual membership. Changes of address for members should be completed online or mailed to the APSA membership office at: APSA, 1527 New Hampshire Avenue, N.W., Washington, DC 20036.

**Subscription, Publishing, and Advertising Office Address:** Cambridge University Press, 40 West 20th Street, New York, NY 10011-4211; and (outside the United States, Canada, and Mexico) Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, England.

**Subscription Information:** *American Political Science Review* (ISSN 0003-0554) is published quarterly in February, May, August, and November by Cambridge University Press for the American Political Science Association. Annual institutional electronic-only subscription rate (2005) is US\$425 in the United States, Canada, and Mexico, UK£268 elsewhere. Annual institutional print-and-electronic subscription rate (2005) is US\$520 in the United States, Canada, and Mexico, UK£322 elsewhere. Periodicals postage rate paid at New York, NY and additional mailing offices. Institutional subscription orders and address changes should be sent to: *American Political Science Review*, Cambridge University Press, 100 Brook Hill Drive, West Nyack, NY 10994-2133, USA.

Copyright © 2005 by the American Political Science Association.

All rights reserved. No part of this publication may be reproduced, in any form or by any means, electronic, photocopying, or otherwise, without permission in writing from Cambridge University Press. Photocopying information for users in the U.S.A.: The Item-Fee Code for the publication (0003-0554/05 \$9.00 + \$.10) indicates that copying for internal or personal use beyond that permitted by Sec. 107 or 108 of the U.S. Copyright Law is authorized for users duly registered with the Copyright Clearance Center (CCC) Transaction Reporting Service, provided that the appropriate remittance of \$9.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. Specific written permission must be obtained for all other copying.

# American Political Science Review

Volume 99, Number 3, August 2005

## Table of Contents



NOTES FROM THE EDITOR .....	iii
ARTICLES	
Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina <b>Susan C. Stokes</b> .....	315
A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective <b>Tulia G. Falletti</b> .....	327
Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success <b>Bonnie M. Meguid</b> .....	347
Uncovering Evidence of Conditional Party Government: Reassessing Majority Party Influence in Congress and State Legislatures <b>William T. Bianco and Itai Sened</b> .....	361
Racism Revised: Courts, Labor Law, and the Institutional Construction of Racial Animus <b>Paul Frymer</b> .....	373
Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure <b>B. Peter Rosendorff</b> .....	389
Reading Habermas in Anarchy: Multilateral Diplomacy and Global Public Spheres <b>Jennifer Mitzen</b> .....	401
Prenegotiation Public Commitment in Domestic and International Bargaining <b>Bahar Leventoglu and Ahmer Tarar</b> .....	419
Nested Analysis as a Mixed-Method Strategy for Comparative Research <b>Evan S. Lieberman</b> .....	435
FORUM	
No Rest for the Democratic Peace <b>David Kinsella</b> .....	453
Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace <b>Branislav L. Slantchev, Anna Alexandrova, and Erik Gartzke</b> .....	459
Three Pillars of the Liberal Peace <b>Michael W. Doyle</b> .....	463
Explaining the Democratic Peace <b>Sebastian Rosato</b> .....	467

# AMERICAN POLITICAL SCIENCE ASSOCIATION

## Officers

**PRESIDENT**  
Margaret Levi  
University of Washington

**PRESIDENT-ELECT**  
Ira Katznelson  
Columbia University

**VICE-PRESIDENTS**  
Robert Axelrod  
University of Michigan

John A. Garcia  
University of Arizona

Joan C. Tronto  
CUNY Graduate Center and  
Hunter College

**SECRETARY**  
Christine Marie Sierra  
University of New Mexico

**TREASURER**  
Henry E. Brady  
University of California,  
Berkeley

**PROGRAM CO-CHAIRS**  
James D. Johnson  
University of Rochester

Jack Knight  
Washington University,  
St. Louis

Susan C. Stokes  
University of Chicago

**EDITOR, APSR**  
Lee Sigelman  
George Washington University

**EDITOR, PERSPECTIVES**  
Jennifer L. Hochschild  
Harvard University

**EXECUTIVE DIRECTOR**  
Michael A. Brintnall  
American Political Science  
Association

KP 3868

## Council

### 2003-2005

Tony Affigne  
Providence College

John H. Aldrich  
Duke University

Manuel Avalos  
Arizona State University West

Judith A. Baer  
Texas A&M University

Shirley M. Geiger  
Savannah State University

John W. Harbeson  
CUNY-City College of  
New York

Marion E. Orr  
Brown University

Joanna Vecchiarelli Scott  
Eastern Michigan University

### 2004-2006

Lisa Anderson  
Columbia University

Andrew L. Aoki  
Augsburg College

Neta C. Crawford  
Brown University

James Louis Gibson  
Washington University,  
St. Louis

Harvey C. Mansfield  
Harvard University

Rogers M. Smith  
University of Pennsylvania

Aili Tripp  
University of Wisconsin,  
Madison

David Vogel  
University of California,  
Berkeley

## Former Presidents

Frank J. Goodnow  
Albert Shaw  
Frederick N. Judson  
James Bryce  
A. Lawrence Lowell  
Woodrow Wilson  
Simeon E. Baldwin  
Albert Bushnell Hart  
W. W. Willoughby  
John Bassett Moore  
Ernst Freund  
Jesse Macy  
Munroe Smith  
Henry Jones Ford  
Paul S. Reinsch  
Leo S. Rowe  
William A. Dunning  
Harry A. Garfield  
James W. Garner  
Charles E. Merriam  
Charles A. Beard  
William Bennett Munro  
Jesse S. Reeves  
John A. Fairlie  
Benjamin F. Shambaugh

Edward S. Corwin  
William F. Willoughby  
Isidor Loeb  
Walter J. Shepard  
Francis W. Coker  
Arthur N. Holcombe  
Thomas Reed Powell  
Clarence A. Dykstra  
Charles Grove Haines  
Robert C. Brooks  
Frederic A. Ogg  
William Anderson  
Robert E. Cushman  
Leonard D. White  
John Gaus  
Walter F. Dodd  
Arthur W. MacMahon  
Henry R. Spencer  
Quincy Wright  
James K. Pollock  
Peter H. Odegard  
Luther Gulick  
Pendleton Herring  
Ralph J. Bunche  
Charles McKinley

Harold D. Lasswell  
E. E. Schattschneider  
V. O. Key Jr.  
R. Taylor Cole  
Carl B. Swisher  
Emmette S. Redford  
Charles S. Hyneman  
Carl J. Friedrich  
C. Herman Pritchett  
David B. Truman  
Gabriel A. Almond  
Robert A. Dahl  
Merle Fainsod  
David Easton  
Karl W. Deutsch  
Robert E. Lane  
Heinz Eulau  
Robert E. Ward  
Avery Leiserson  
Austin Ranney  
James MacGregor Burns  
Samuel H. Beer  
John C. Wahlke  
Leon D. Epstein  
Warren E. Miller

Charles E. Lindblom  
Seymour Martin Lipset  
William H. Riker  
Philip E. Converse  
Richard F. Fenno Jr.  
Aaron B. Wildavsky  
Samuel P. Huntington  
Kenneth N. Waltz  
Lucian W. Pye  
Judith N. Shklar  
Theodore J. Lowi  
James Q. Wilson  
Lucius J. Barker  
Charles O. Jones  
Sidney Verba  
Arend Lijphart  
Elinor Ostrom  
M. Kent Jennings  
Matthew Holden Jr.  
Robert O. Keohane  
Robert Jervis  
Robert D. Putnam  
Theda Skocpol  
Susanne Hoerber Rudolph



## Notes from the Editor

### In This Issue<sup>1</sup>

Is political science the real dismal discipline? One might think so, given the head-shaking, hand-wringing, and tut-tutting for which political scientists are responsible during every election cycle. Too few citizens, we lament, take the trouble to vote, and too many of those who do vote base their decisions on superficial or whimsical grounds. The unease we feel as professionals-cum-citizens over the distance between the noble idea of elections-in-theory and the sorry conduct of elections-in-practice has a long pedigree. In the first century A.D., Juvenal decried the tendency of imperial politicians to sweep serious policy issues under the rug by satiating the populace with *panem et circenses*. Colonial-era British politicians also courted votes with food and drink. The famous 1757 painting "Canvassing for Votes" by William Hogarth depicts vote-seekers gaining electoral support based upon their skills as genial hosts, not policy advocates. In many American cities, elections have long been notoriously corrupt, the classic case being New York's Tammany Hall and its ethos of "I seen my opportunities and I took 'em." Today, as fledgling democracies around the world are holding elections, they are experiencing many of the forms of electoral corruption and graft that have become so familiar in more established democracies, and undoubtedly they are devising some new forms as well.

Argentina cast off military rule just two decades ago. Susan C. Stokes demonstrates that parties there, as elsewhere, use material inducements and social pressures to try to gain support on Election Day. In "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina," Stokes uses a broad range of methodological tools to analyze the electoral tactics of political machines. Her analysis should be of particular interest to both comparativists and Americanists, and it should serve more generally as a reminder of both the promise and pitfalls of electoral democracy.

Argentina reappears in Tulia G. Falleti's "A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective." Decentralization is often seen as empowering subnational leaders at the expense of the central government. Falleti argues instead that decentralization has administrative, fiscal, and political dimensions, the combination of which does not inevitably lead to greater subnational power. Rather, the interplay of sequence and interlevel interests determines the course and consequences of decentralization. Local leaders prefer autonomy, money, and then responsibility, but a different ordering could leave subnational governments burdened with unfunded mandates. Based on fresh ideas and revealing interviews with local officials in several Latin American countries,

Falleti's study is likely to lead to a reconsideration of widely accepted ideas about decentralization and its effects.

In many established democracies, greens, ultra-nationalists, and other non-"mainstream" parties, once mere footnotes in electoral politics, are "playing with the big boys now." Bonnie M. Meguid examines the emergence and performance of new, single-issue, or "niche" parties in "Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success." Existing explanations, Meguid argues, pay insufficient heed to the mainstream parties' strategic responses to the threat that niche parties pose to their hegemony. Accordingly, Meguid develops a modified spatial model and uses it to assess the impact of mainstream parties' strategies on the electoral performance of niche parties in 17 Western European countries.

Notwithstanding Vince Lombardi's dictum that "Winning isn't everything—it's the only thing," winning elections is only the first hurdle for political parties. The task of governing remains. But do parties really matter insofar as governing is concerned, or—at least in the American context—is party just a label? This question divides students of congressional politics. Much debate has taken place at the theoretical level, with each side ceding little ground to the other. In "Uncovering Evidence of Conditional Party Government: Reassessing Majority Party Influence in Congress and State Legislatures," William T. Bianco and Itai Sened take the discussion to the next level by evaluating expectations drawn from the competing theories. Drawing on data from several sessions of Congress and several state legislatures, Bianco and Sened conclude that party leaders are more like chessmasters than cat-herders, often using their influence to set the agenda and to structure outcomes in favor of their parties' interests. These findings constitute an important addition to our understanding of the role of parties in legislatures and provide a foundation for additional research.

Issues involving race and ethnicity are never far from center stage in the play of American politics. Paul Frymer takes contemporary explanations of racism to task for emphasizing individual-level psychological factors at the expense of institutional ones. Making innovative use of data from the National Labor Relations Board's handling of cases of alleged racism in union elections, Frymer explores how rules, institutions, and politics can contribute to individual acts of racism. Both general readers and specialists in the politics of race and ethnicity will find much of interest in "Racism Revisited: Courts, Labor Law and the Institutional Construction of Racial Animus."

Other than their shared focus on international relations and negotiations, the next three articles in this issue may seem to have little in common. Each of them,

<sup>1</sup> Drafted by APSR editorial assistant Lee Michael.

however, demonstrates that a human touch is often necessary to navigate safely through various diplomatic pitfalls and obstacles.

In an era of globalization and free trade, governments are often conflicted about honoring international trade agreements, lest they be viewed as insincere abroad, without angering citizens anxious about job security, lest they risk defeat at the polls. This Putnamesque insight underlies B. Peter Rosendorff's "Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure." Rosendorff argues that the World Trade Organization's dispute settlement procedure enables states to have it both ways by suspending their obligations temporarily during periods of increased domestic pressure for protectionism. Because this analysis assesses the balance between rigidity and stability in the design of international institutions, is likely to resonate across a wide readership, ranging from scholars concerned with institutional design to those concerned more generally with the relationship between the international and domestic arenas and the effect that this intersection has on policy outcomes.

People behave differently—often "better"—when they know they are being watched. That, according to Jennifer Mitzen, is particularly true for diplomats who must explain their country's positions to other diplomats across the negotiating table; the simple act of talking things out in a visible, public forum can "refine and enlarge" views of allies, adversaries, and even enemies. Mitzen's "Reading Habermas in Anarchy: Multilateral Diplomacy and Global Public Spheres" contributes significantly to international relations scholarship by treating horizontal discourse between states as a public sphere capable of legitimating state action and mitigating anarchy, and broadens the theoretical foundation for scholars interested in a wide range of topics, including the security dilemma, global governance, the democratic peace, and discourse theory.

Before trying to scale a high fence, it can help to throw something valuable over the top first; that should enhance the motivation to succeed. Political leaders employ a similar logic when they publicly predict negotiating successes in hopes of precluding unwanted compromises or concessions, argue Bahar Leventoglu and Ahmer Tarar in "Prenegotiation Public Commitment in Domestic and International Bargaining." The structure of the bargaining situation provides incentives to overstate one's goals, which, in turn, should maximize one's potential gains. The danger is that when all parties at the table use this tactic, the likelihood of deadlock is greatly increased. Leventoglu and Tarar's analysis provides a formal proof of the common wisdom that agreements and compromises are best forged in secret, as Middle East peace negotiators, constitutional convention delegates, and sequestered cardinals can all attest.

Large-N or small? Both approaches to comparative research have their advantages and their limitations. In "Nested Analysis as a Mixed-Method Strategy for Comparative Research," Evan S. Lieberman offers a much-needed guide for combining the two approaches

in a single research design, in the form of a nested analysis. A mixed strategy of using the large-N approach in case selection and casual inference and the small-N approach to inform measurement and model specification can, Lieberman contends, greatly enhance the methodological quality of research and thereby bolster the validity and reliability of research results.

Few predictions have ever seemed safer than one that was issued in our November 2003 "Notes from the Editor," to the effect that Sebastian Rosato's "The Flawed Logic of Democratic Peace Theory" would be "sure to stir controversy." The trio of responses to Rosato's article that appear in the "Forum" section of the current issue indicate the great interest and high feelings that surround democratic peace theory. The controversy turns less on the empirical regularity of peace between democracies itself than on the explanation for this phenomenon. Is there something inherently different about the *modus operandi* of democracies, as democratic peace theory advocates contend, or has a *pax Americana* imposed order and stability over Western Europe and the New World during the post-World War II era, as realists like Rosato argue?

In "No Rest for the Democratic Peace", David Kinsella argues that because democratic peace theory is dyadic in its logic, not monadic, much of Rosato's monadically-based analysis is off-target. Branislav L. Slantchev, Anna Alexandrova, and Erik Gartzke, in "Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace," find in Rosato's analysis an insufficient appreciation of the probabilistic nature of democratic peace theory, and go on to raise concerns about the impact of selection bias on the substantive results that he reports. Returning in "Three Pillars of the Liberal Peace" to the Kantian basis of democratic peace theory, Michael W. Doyle reminds all involved that republican representation, support for human rights, and transnational interdependence work to produce democratic peace only jointly.

Responding to these critiques in "Explaining the Democratic Peace?," Rosato stands by his original points. To Kinsella, Rosato concedes that the empirical regularity on which democratic peace theory is based is dyadic, but emphasizes that the six original logics he identified are monadic. To the methodological concerns of Slantchev, Alexandrova, and Gartzke, Rosato does not disagree that the theory is probabilistic, but sees it as failing even when understood as such; he also argues that new evidence on accountability makes the selection bias charge unconvincing. Finally, Rosato concurs with Doyle that Kantian democracies will rarely go to war but sees their co-pacifism as having little to do with democracy.

This four-sided exchange concludes the discussion insofar as the *APSR* is concerned, but another safe prediction is that it will not conclude the discussion overall. As debate and research continue on the root causes of war and peace, we hope that this "Forum" exchange will play a useful role in clarifying the remaining theoretical, conceptual, and methodological issues.



## INSTRUCTIONS TO CONTRIBUTORS

### General Considerations

The *APSR* strives to publish scholarly research of exceptional merit, focusing on important issues and demonstrating the highest standards of excellence in conceptualization, exposition, methodology, and craftsmanship. Because the *APSR* reaches a diverse audience of scholars and practitioners, authors must demonstrate how their analysis illuminates a significant research problem, or answers an important research question, of general interest in political science. For the same reason, authors must strive for a presentation that will be understandable to as many scholars as possible, consistent with the nature of their material.

The *APSR* publishes original work. Therefore, authors should not submit articles containing tables, figures, or substantial amounts of text that have already been published or are forthcoming in other places, or that have been included in other manuscripts submitted for review to book publishers or periodicals (including on-line journals). In many such cases, subsequent publication of this material would violate the copyright of the other publisher. The *APSR* also does not consider papers that are currently under review by other journals or duplicate or overlap with parts of larger manuscripts that have been submitted to other publishers (including publishers of both books and periodicals). Submission of manuscripts substantially similar to those submitted or published elsewhere, or as part of a book or other larger work, is also strongly discouraged. If you have any questions about whether these policies apply in your particular case, you should discuss any such publications related to a submission in a cover letter to the Editor. You should also notify the Editor of any related submissions to other publishers, whether for book or periodical publication, that occur while a manuscript is under review by the *APSR* and which would fall within the scope of this policy. The Editor may request copies of related publications.

If your manuscript contains quantitative evidence and analysis, you should describe your procedures in sufficient detail to permit reviewers to understand and evaluate what has been done and, in the event that the article is accepted for publication, to permit other scholars to carry out similar analyses on other data sets. For example, for surveys, at the least, sampling procedures, response rates, and question wordings should be given; you should calculate response rates according to one of the standard formulas given by the American Association for Public Opinion Research, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (Ann Arbor, MI: AAPOR, 2000). This document is available on the Internet at <[http://www.aapor.org/default.asp?page=survey\\_methods/standards\\_and\\_best\\_practices/standard\\_definitions](http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/standard_definitions)>. For experiments, provide full descriptions of experimental protocols, methods of subject recruitment and selection, subject payments and debriefing procedures, and so on. Articles should be self-contained, so you should not simply refer read-

ers to other publications for descriptions of these basic research procedures.

Please indicate variables included in statistical analyses by capitalizing the first letter in the variable name and italicizing the entire variable name the first time each is mentioned in the text. You should also use the same names for variables in text and tables and, wherever possible, should avoid the use of acronyms and computer abbreviations when discussing variables in the text. All variables appearing in tables should have been mentioned in the text and the reason for their inclusion discussed.

As part of the review process, you may be asked to submit additional documentation if procedures are not sufficiently clear; the review process works most efficiently if such information is given in the initial submission. If you advise readers that additional information is available, you should submit printed copies of that information with the manuscript. If the amount of this supplementary information is extensive, please inquire about alternate procedures.

The *APSR* uses a double-blind review process. You should follow the guidelines for preparing anonymous copies in the Specific Procedures section below.

Manuscripts that are largely or entirely critiques or commentaries on previously published *APSR* articles will be reviewed using the same general procedures as for other manuscripts, with one exception. In addition to the usual number of reviewers, such manuscripts will also be sent to the scholar(s) whose work is being criticized, in the same anonymous form that they are sent to reviewers. Comments from the original author(s) to the Editor will be invited as a supplement to the advice of reviewers. This notice to the original author(s) is intended (1) to encourage review of the details of analyses or research procedures that might escape the notice of disinterested reviewers; (2) to enable prompt publication of critiques by supplying criticized authors with early notice of their existence and, therefore, more adequate time to reply; and (3) as a courtesy to criticized authors. If you submit such a manuscript, you should therefore send as many additional copies of their manuscripts as will be required for this purpose.

Manuscripts being submitted for publication should be sent to Lee Sigelman, Editor, *American Political Science Review*, Department of Political Science, The George Washington University, Washington, DC 20052. Correspondence concerning manuscripts under review may be sent to the same address or e-mailed to [apsr@gwu.edu](mailto:apsr@gwu.edu).

### Manuscript Formatting

Manuscripts should not be longer than 45 pages including text, all tables and figures, notes, references, and appendices. This page size guideline is based on the U.S. standard 8.5 × 11-inch paper; if you are submitting a manuscript printed on longer paper, you must adjust accordingly. The font size must be at least 11 points for all parts of the paper, including notes and references. The entire paper, including notes and references, must be double-spaced, with the sole exception of tables

for which double-spacing would require a second page otherwise not needed. All pages should be numbered in one sequence, and text should be formatted using a normal single column no wider than 6.5 inches, as is typical for manuscripts (rather than the double-column format of the published version of the *APSR*), and printed on one side of the page only. Include an abstract of no more than 150 words. The *APSR* style of embedded citations should be used, and there must be a separate list of references at the end of the manuscript. Do not use notes for simple citations. These specifications are designed to make it easier for reviewers to read and evaluate papers. Papers not adhering to these guidelines are subject to being rejected without review.

For submission and review purposes, you may place footnotes at the bottom of the pages instead of using endnotes, and you may locate tables and figures (on separate pages and only one to a page) approximately where they fall in the text. However, manuscripts accepted for publication must be submitted with endnotes, and with tables and figures on separate pages at the back of the manuscript with standard indications of text placement, e.g., [Table 3 about here]. In deciding how to format your initial submission, please consider the necessity of making these changes if your paper is accepted. If your paper is accepted for publication, you will also be required to submit camera-ready copy of graphs or other types of figures. Instructions will be provided.

For specific formatting style of citations and references, please refer to articles in the most recent issue of the *APSR*. For unusual style or formatting issues, you should consult the latest edition of *The Chicago Manual of Style*. For review purposes, citations and references need not be in specific *APSR* format, although some generally accepted format should be used, and all citation and reference information should be provided.

## Specific Procedures

Please follow these specific procedures for submission:

1. You are invited to submit a list of scholars who would be appropriate reviewers of your manuscript. The Editor will refer to this list in selecting reviewers, though there obviously can be no guarantee that those you suggest will actually be chosen. Do not list anyone who has already commented on your paper or an earlier version of it, or any of your current or recent collaborators, institutional colleagues, mentors, students, or close friends.
2. Submit five copies of manuscripts *and* a diskette or CD containing a pdf file of the anonymous version of the manuscript. If you cannot save the manuscript as a pdf, just send in the diskette or CD with the word-processed version. Please ensure that the paper and diskette or CD versions you submit are identical; the diskette or CD version should be of the **anonymous**

copy (see below). Please review all pages of all copies to make sure that all copies contain all tables, figures, appendices, and bibliography mentioned in the manuscript and that all pages are legible. Label the diskette or CD clearly with the (first) author's name and the title of the manuscript (in abridged form if need be), and identify the word processing program and operating system. If you are unable to create a diskette or CD, please note this in your submission, and you will be asked to e-mail the appropriate file.

3. To comply with the *APSR*'s procedure of double-blind peer reviews, only one of the five copies submitted should be fully identified as to authorship and four should be in anonymous format.
4. For anonymous copies, if it is important to the development of the paper that your previous publications be cited, please do this in a way that does not make the authorship of the submitted paper obvious. This is usually most easily accomplished by referring to yourself in the third person and including normal references to the work cited in the list of references. In no circumstances should your prior publications be included in the bibliography in their normal alphabetical location but with your name deleted. Assuming that text references to your previous work are in the third person, you should include full citations as usual in the bibliography. Please discuss the use of other procedures to render manuscripts anonymous with the Editor prior to submission. You should not thank colleagues in notes or elsewhere in the body of the paper or mention institution names, web page addresses, or other potentially identifying information. *All acknowledgments must appear on the title page of the identified copy only.* Manuscripts that are judged not anonymous will not be reviewed.
5. The first page of the four anonymous copies should contain only the title and an abstract of no more than 150 words. The first page of the identified copy should contain (a) the name, academic rank, institutional affiliation, and contact information (mailing address, telephone, fax, e-mail address) for all authors; (b) in the case of multiple authors, an indication of the author who will receive correspondence; (c) any relevant citations to your previous work that have been omitted from the anonymous copies; and (d) acknowledgments, including the names of anyone who has provided comments on the manuscript. If the identified copy contains any unique references or is worded differently in any way, please mark this copy with "Contains author citations" at the top of the first page.

No copies of submitted manuscripts can be returned.



## ELECTRONIC ACCESS TO THE *APSR*

Back issues of the *APSR* are available in several electronic formats and through several vendors. Except for the last three years (as an annually "moving wall"), back issues of the *APSR* beginning with Volume 1, Number 1 (November 1906), are available on-line through JSTOR (<http://www.jstor.org/>). At present, JSTOR's complete journal collection is available only via institutional subscription, e.g., through many college and university libraries. For APSA members who do not have access to an institutional subscription to JSTOR, individual subscriptions to its *APSR* content are available. Please contact Member Services at APSA for further information, including annual subscription fees.

Individual members of the American Political Science Association can access recent issues of the *APSR* and *PS* through the APSA website ([www.apsanet.org](http://www.apsanet.org)) with their username and password. Individual non-member access to the online edition will also be available, but only through institutions that hold either a print-plus-electronic subscription or an electronic-only subscription, provided the institution has registered and activated its online subscription.

Full text access to current issues of both the *APSR* and *PS* is also available on-line by library subscription from a number of database vendors. Currently, these include University Microfilms Inc. (UMI) (via its CD-ROMs General Periodicals Online and Social Science Index and the on-line database ProQuest Direct), Online Computer Library Center (OCLC) (through its on-line database First Search as well as on CD-ROMs and magnetic tape), and the Information Access Company (IAC) (through its products Expanded Academic Index, InfoTrac, and several on-line services [see below]). Others may be added from time to time.

The *APSR* is also available on databases through six online services: Datastar (Datastar), Business Library (Dow Jones), Cognito (IAC), Encarta Online Library (IAC), IAC Business (Dialog), and Newsearch (Dialog).

The editorial office of the *APSR* is not involved in the subscription process to either JSTOR for back issues or the other vendors for current issues. Please contact APSA, your reference librarian, or the database vendor for further information about availability.

## BOOK REVIEWS

The *APSR* no longer contains book reviews. As of 2003, book reviews have moved to *Perspectives on Politics*. All books for review should be sent to the *Perspectives on Politics* Book Review Editor, Jeffrey C. Isaac. The address is Professor Jeffrey C. Isaac, Review Editor, *Perspectives on Politics*, Department of Political Science, Woodburn Hall, 1100 E. 7th St., Indiana University, Bloomington, IN 47405-7110. E-mail: [isaac@indiana.edu](mailto:isaac@indiana.edu).

If you are the author of a book you wish to be considered for review, please ask your publisher to

send a copy to the *Perspectives on Politics* Book Review Editors per the mailing instructions above. If you are interested in reviewing books for *Perspectives on Politics*, please send your vita to the Book Review Editors; you should not ask to review a specific book.

## OTHER CORRESPONDENCE

The American Political Science Association's address, telephone, and fax are 1527 New Hampshire Avenue, N.W., Washington, D.C. 20036, (202) 483-2512 (voice), and (202) 483-2657 (fax). E-mail: [apsa@apsanet.org](mailto:apsa@apsanet.org). Please direct correspondence as follows.

Information, including news and notes, for *PS*:

Dr. Robert J-P. Hauck, Editor, *PS*  
E-mail: [rhauck@apsanet.org](mailto:rhauck@apsanet.org)

Circulation and subscription correspondence (domestic claims for nonreceipt of issues must be made within four months of the month of publication; overseas claims, within eight months):

Sean Twombly,  
Director of Member Services  
E-mail: [membership@apsanet.org](mailto:membership@apsanet.org)

Reprint permissions:  
E-mail: [Rights@cambridge.org](mailto:Rights@cambridge.org)

Advertising information and rates:

Advertising Coordinator,  
Cambridge University Press  
E-mail: [advertising@apsanet.org](mailto:advertising@apsanet.org)

## EXPEDITING REQUESTS FOR COPYING *APSR* AND *PS* ARTICLES FOR CLASS USE AND OTHER PURPOSES

### Class Use

The Comprehensive Publisher Photocopy Agreement between APSA and the Copyright Clearance Center (CCC) permits bookstores and copy centers to receive expedited clearance to copy articles from the *APSR* and *PS* in compliance with the Association's policies and applicable fees. The general fee for articles is 75 cents per copy. However, current Association policy levies no fee for the first 10 copies of a printed article, whether in course packs or on reserve. Smaller classes that rely heavily on articles (i.e., upper-level undergraduate and graduate classes) can take advantage of this provision, and faculty ordering 10 or fewer course packs should bring it to the attention of course pack providers. APSA policy also permits free use of the electronic library reserve, with no limit on the number

of students who can access the electronic reserve. Both large and small classes that rely on these articles can take advantage of this provision. The CCC's address, telephone, and fax are 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400 (voice), and (978) 750-4474 (fax). This agreement pertains only to the reproduction and distribution of APSA materials as hard copies (e.g., photocopies, microfilm, and microfiche).

The Association of American Publishers (AAP) has created a standardized form for college faculty to submit to a copy center or bookstore to request copyrighted material for course packs. The form is available through the CCC, which will handle copyright permissions.

APSA also has a separate agreement pertaining to CCC's Academic E-Reserve Service. This agreement allows electronic access for students and instructors of a designated class at a designated institution for a specified article or set of articles in electronic format. Access is by password for the duration of a class.

Please contact your librarian, the CCC, or the APSA Reprints Department for further information.

### **APSR Authors**

If you are the author of an *APSR* article, you may use your article in course packs or other printed materials without payment of royalty fees and you may post it at personal or institutional web sites as long as the APSA copyright notice is included.

### **Other Uses of APSA-Copyrighted Materials**

For any further copyright issues, please contact the APSA Reprints Department.

### **INDEXING**

Articles appearing in the *APSR* before June 1953 were indexed in *The Reader's Guide to Periodical Literature*. Current issues are indexed in *ABC Pol Sci*; *America, History and Life 1954-*; Book Review Index; Current Contents: Social and Behavioral Sciences; Econ-Lit; *Energy Information Abstracts*; Environmental Abstracts; Historical Abstracts; Index of Economic Articles; Information Service Bulletin; International Index; International Political Science Abstracts; the *Journal of Economic Literature*; Periodical Abstracts; Public Affairs; Public Affairs Information Service International Recently Published Articles; Reference Sources; Social Sciences and Humanities Index; Social Sciences Index; Social Work Research and Abstracts; and Writings on American History. Some of these sources may be available in electronic form through local public or educational libraries. Microfilm of the *APSR*, beginning with Volume 1, and the index of the *APSR* through 1969 are available through University Microfilms Inc., 300 North Zeeb Road, Ann Arbor, MI 48106 ([www.umi.com](http://www.umi.com)). The Cumulative Index to the *American Political Science Review*, Volumes 63 to 89: 1969-95, is available through the APSA.



# Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina

SUSAN C. STOKES *Yale University*

*Political machines (or clientelist parties) mobilize electoral support by trading particularistic benefits to voters in exchange for their votes. But if the secret ballot hides voters' actions from the machine, voters are able to renege, accepting benefits and then voting as they choose. To explain how machine politics works, I observe that machines use their deep insertion into voters' social networks to try to circumvent the secret ballot and infer individuals' votes. When parties influence how people vote by threatening to punish them for voting for another party, I call this perverse accountability. I analyze the strategic interaction between machines and voters as an iterated prisoners' dilemma game with one-sided uncertainty. The game generates hypotheses about the impact of the machine's capacity to monitor voters, and of voters' incomes and ideological stances, on the effectiveness of machine politics. I test these hypotheses with data from Argentina.*

Thirty-five years ago, James Scott (1969) observed that political life of contemporary new nations bore a strong resemblance to the machine politics of the United States in earlier eras. The patronage, particularism, and graft endemic to the Philippines or Malaysia in the postwar decades recalled, for Scott, the Tweed machine in nineteenth-century New York or the Dawson machine in twentieth-century Chicago. Much has happened in the third of a century since Scott outlined "the contours and dynamics of the 'machine model' in comparative perspective" (1143). Many of the new nations that occupied his analysis have undergone transitions to electoral democracy; yet politics in these systems often remains particularistic, clientelistic, and corrupt. We therefore have a larger sample of countries, and a richer experience on which to draw, to understand the contours and dynamics of the machine. The historiography of the U.S. political machine has also grown, as have historical studies of patronage and vote buying in the history of today's advanced European democracies (see, e.g., Piattoni 2001). Finally, a formal literature on redistributive politics has developed, one in which the political machine plays a central role.

Yet the formal literature on the political machine leaves some crucial questions unanswered. Chief among them: How does the machine keep voters from reneging on the implicit deal whereby the machine distributes goods and the recipient votes for the machine? If voters can renege, then machines should not waste scarce resources on them and clientelist politics breaks down. The question is the more pressing, given that many of the societies in which we find active political

machines also have the secret ballot. Political machines did not disappear in the United States after the introduction of the Australian ballot in most U.S. states at the end of the nineteenth century.<sup>1</sup> And clientelism flourishes in countries from Mexico (Fox 1994) to Italy (Chubb 1982) to Bulgaria (Kitschelt et al. 1999), all of which have the ballot.

Assuming that machines can overcome the problem of their clients' reneging, what kinds of voters will they target? Scattered through the qualitative literature is evidence that poor voters are the targets of machines (see, e.g., Chubb 1982; Wilson and Banfield 1963). Formal treatments agree, citing diminishing marginal utility of income as the reason why particularistic benefits generate more votes among the poor than the rich (Calvo and Murillo 2004; Dixit and Londregan 1996).

Yet in the societies where clientelistic parties or machines are active, not all poor voters receive benefits. Limited resources force political machines to choose among poor voters. Machine operatives everywhere face a version of the dilemma that an Argentine Peronist explains. About 40 voters live in her neighborhood, and her responsibility is to get them to the polls and get them to vote for her party. But the party gives her only 10 bags of food to distribute, "ten little bags," she laments, "nothing more."<sup>2</sup> How does she, and machine operatives like her in systems around the world, decide who among her neighbors shall and who shall not receive handouts?

The formal literature answers this question by saying that machines target *core constituents*. But if these constituents are ideologically committed to the machine, is it not wasting resources if it distributes rewards to them? Would it not do better by distributing rewards to the uncommitted or even to those who, on ideological

Susan C. Stokes is the John S. Saden Professor of Political Science, Department of Political Science, P.O. Box 208301, Yale University, New Haven, Ct 06520-8301. (susan.stokes@yale.edu).

I thank Valeria Brusco, John Carey, Matt Cleary, Avinash Dixit, Jeff Grynawski, John Londregan, Scott Mainwaring, Roger Myerson, Luis Fernando Medina, Mario Navarro, Marcelo Nazareno, Steve Pincus, Duncan Snidal, and three anonymous *APSR* reviewers for excellent comments. I am especially indebted to Michael Wallerstein. Research was supported by National Science Foundation research Grant SES-0241958 and by the John Simon Guggenheim Memorial Foundation.

<sup>1</sup> The Australian ballot is one in that it is produced by governments or neutral election authorities (rather than by political parties), distributed through guarded channels on or close to election day, and that lists all parties or candidates for an office in a single format.

<sup>2</sup> Interview conducted by Valeria Brusco, Susan Stokes, and Gloria Trocillo, in Villa Mercedes, Argentina, July 2003; my translation. This and all subsequent translations from the Spanish are by the author.

grounds, oppose the party? The selection method of our Argentine operative is to help “the people who complain the most, the ones who say, ‘What are you going to give me?’ I pick them up [to take them to the polls] and after I take them they say, ‘Aren’t there any bags of food?’” Her words hint at a logic in which machines give private handouts not to die-hard supporters but to people whose future support is in doubt. The analysis in this article helps make sense of her explanation.

Far from being just a Latin American problem, or a problem that advanced democracies have completely overcome, vote buying, clientelism, and machine politics are blights on many democracies around the world, even today. Prosecutors in 2004 accused a candidate for a district judgeship in Eastern Kentucky of giving \$50 checks to voters, implicitly in return for their support.<sup>3</sup> Journalists reported, also in 2004, that an elderly hospital patient in Ukraine confessed to his son that he had voted for the official presidential candidate, Viktor Yanukovich, rather than for the opposition candidate, Viktor Yushenko. He had planned to support Yushenko but switched his vote after a nurse at the hospital promised him a wheelchair if he switched.<sup>4</sup>

These practices make a mockery of democratic accountability. Democratic accountability usually means that voters know, or can make good inferences about, what parties have done in office and reward or punish them conditional on these actions. But when parties know, or can make good inferences about, what individual voters have done in the voting booth and reward or punish them conditional on these actions, this is *perverse accountability*. We usually think of accountability in democratic systems as a good thing: it means that voters can keep elected officials from misbehaving and pressure governments to be more responsive to voters. But perverse accountability is bad for democracy: it reduces the pressure on governments to perform well and to provide public goods, keeps voters from using elections to express their policy preferences, and undermines voter autonomy (see Karlan 1994; Kochin and Kochin 1998; O'Donnell 1996; Stokes 2004). To overcome perverse accountability, we need first to understand how machine politics works. This article begins to build such an understanding.

## STATIC MODELS OF REDISTRIBUTIVE POLITICS AND THE COMMITMENT PROBLEM

In some of our leading formal models of redistributive politics, the political machine plays a large role. Dixit and Londregan (1996) model the strategies of two parties as they attempt to mobilize groups of voters, who care both about consumption and about ideology. Parties tax some voters and redistribute to others. When both parties are equally able to deliver resources to

every group, the parties deploy tactical rewards to compete for the same groups of *swing* voters—groups with a relatively large number of moderate voters who are ideologically indifferent between the two parties. But when one party has an especially close link to a group of voters, then the party will target this *core constituency*. Dixit and Londregan write that core constituents are ones

whom [the party] understands well . . . A party's core constituencies need not prefer its issue positions. It is the party's advantage over its competitors at swaying voters in a group with offers of particularistic benefits that makes the group core (1986, 1134). . . . The key to the electoral strategies of the urban political machines was their ability to provide “personal services” to their core constituents at a lower cost than could their competitors. They did this by knowing their constituents (1147).

For Cox and McCubbins (1986), the crucial feature of the machine–core constituent link is that the party is more certain about how core groups will respond to rewards than it is about other groups. The party is more certain because “core supporters . . . are well-known quantities. The candidate is in frequent and intensive contact with them and has relatively precise and accurate ideas about how they will react” (1986, 378–9).

The problem with both pairs of authors' models is that they don't deal adequately with commitment problems. Both assume by caveat that the party won't renege on its offer of particularistic rewards once it's won the election.<sup>5</sup> And they don't deal adequately with the fact that a voter, once in the voting booth, can also renege by voting his or her conscience or preference, ignoring the reward he or she received. When we translate these authors' models into one-shot strategic interactions between party operatives and voters, redistributive politics does not happen. (For reasons of space, I do not analyze such games here.) The operative doesn't give a reward, and the possibility of a reward doesn't change the voter's vote. This commitment problem looms not only over the relation between machines and core constituents but also over the one between parties and swing voters: the party's dominant strategy is to renege, and the voter's is to vote for the party it prefers on ideological or programmatic grounds, not the one that deployed tactical rewards.

## A DYNAMIC MODEL OF MACHINE POLITICS

### Assumptions

A way to deal with these commitment problems is to place the machine–voter interaction in a dynamic context. To model the interaction between machine operatives and voters as a repeated game, we have to make certain assumptions. First, we have to assume that parties can monitor individual voters' actions and

<sup>3</sup> *The New York Times*, August 29, 2004

<sup>4</sup> “Ukrainian Campaigns Gear Up for Presidential Re-Vote,” Emily Harris, December 7, 2004, [www.npr.org](http://www.npr.org)

<sup>5</sup> Aware that parties in their model suffer from a commitment problem, Cox and McCubbins simply add an assumption “that candidates, once elected, carry out their promises” (1986, 373).



condition rewards on their inferred votes. Second, we have to assume that both sides perceive the interaction as ongoing indefinitely into the future.

The assumption that machines can hold voters accountable, that they can monitor individuals' votes (even if imperfectly) and make rewards contingent on the voter's support, departs from the implicit assumption of redistributive theorists. They assume that a member of a favored group will receive private rewards whether or not he votes for the party; individual voters are anonymous and therefore free from the party's retribution should they defect. The premise that voting is a private and anonymous act may have discouraged formal theorists from modeling these interactions as repeated games; repeated games generally rely on each player being able to observe the actions of the other in the previous round. The assumption that voting is anonymous is appropriate in most advanced democracies, but not necessarily in the historical context of political machines or in many new democracies today.<sup>6</sup>

There are two kinds of private information about the voter that are useful to the party: his actions—which party he votes for—and his type—his partisan predisposition in relation to the two parties.<sup>7</sup> Machines are good at gathering information about voters' actions and types. Indeed, formal theorists have identified features of the machine that makes it good at discerning what people need and delivering it to them efficiently, but these same features *also make it good at discerning individuals' likely votes*.

Certain voting technologies allow parties to monitor individuals' votes. The recent historiography of U.S. machines deepens our appreciation of these technologies. Until the introduction of the Australian ballot in the United States, in most states in 1891, parties produced "ticket" or "coupon" ballots, ones that listed only their candidates. To monitor which party's ballot the voter was using, parties printed ballots on paper of different weights or colors. Voters deposited the ballot directly in the ballot box, under the watchful eye of party operatives, without first concealing it in an envelope (for descriptions, see Keyssar 2000; Reynolds 1988). Reynolds (1980, 193) reports that New Jersey's early automatic voting machines, introduced in 1890, made clicking noises that allowed party officials standing nearby to detect the voter's selection. And operatives from the Philadelphia Republican party in the late nineteenth and early twentieth century offered to fill out ballots on voters' behalf (McCaffery 1993).

Voting practices and technologies undermine the anonymity of the vote in contemporary developing democracies, as they did in U.S. machine cities, even where the Australian ballot is in use and where voting is, in a narrow sense, secret. In her description of contemporary India as a "patronage democracy," Chandra

(2004) notes that parties designate polling agents to observe the progress of voting. Polling agents are "usually men from the village itself, or from close by, who know the identity of each voter. While they do not witness the actual vote, they know who shows up to vote and can report on turnout figures" (139). Chandra reports that Indian parties could undermine voters' anonymity by emptying boxes and counting the returns at frequent intervals over the course of an election day. (An electoral reform in 1994 outlawed the practice.) To cite another example, in the 2003 Russian Duma elections, international observers reported "significant problems relating to the secrecy of the vote, with open voting in 30% ... of polling stations ... polling officials and party observers were seen to be actively encouraging persons to vote outside of polling booths" (Organization for Security and Cooperation in Europe 2003).

Certain party-organizational structures allow parties to discern individual voters' types—their predisposition for or against the machine. The typical political machine (or clientelist party) is bottom-heavy, decentralized, and relies on an army of grassroots militants. Voters in today's democracies in the developing world are frequently geographically immobile, living in neighborhoods where they grew up and where family members and close acquaintances live. Some of these familiar neighbors work as operatives for political parties. They therefore know much about an individual that shapes his partisan attachments: his job, associational membership, parents' ideological inclinations, and public statements about parties and policies. It is also hard for voters to dissemble before people they've known all their lives: as one grassroots party organizer in Argentina explained, you know if a neighbor voted against your party if he can't look you in the eye on election day.

Information about individual voters' partisan predispositions helps the machine make inferences about how individuals vote and whether they are good candidates for vote buying. For instance, the model in the next section shows that voters who are predisposed in favor of the machine on partisan or programmatic grounds cannot credibly threaten to punish their favored party if it withholds rewards. Therefore the party should not waste rewards on them. The model also shows that voters who are strongly opposed to the machine will not trade their votes for rewards. A machine can compensate, to some degree, for an effective secret ballot if it can distinguish strong opponents from people who oppose it more moderately, or strong loyalists from people who are indifferent about whom to vote for.

Argentina, the country from which I present evidence, combines a balloting system that gives parties greater control over voters than does the Australian ballot, a social structure of reduced anonymity, especially among the poor, and party organizations that help parties monitor voters. These features contribute to a widespread perception among Argentine voters and party operatives that voting is a less than fully anonymous act. As one grassroots party organizer explained, "Anyone who's militating in the streets, you

<sup>6</sup> It is not always appropriate for advanced democracies. In contemporary Spain, voters retrieve sheets containing party lists from an open table at their polling place. They can retreat into an enclosed booth to cast their ballot. But they are not required to vote in secret and many vote in the open.

<sup>7</sup> In the models that follow, I assume two-party competition, as do the theorists of redistributive politics discussed earlier.

know who's with you and who's not with you."<sup>8</sup> In a survey conducted in four Argentine provinces in July–August 2003, respondents were asked, "Even though the vote is secret, do you believe that party operatives can find out how a person in your neighborhood has voted?" Despite a technically secret ballot, 37% of the sample responded that party operatives can find out, 51% that they cannot, and the remaining 12% didn't know (total sample size: 2,000).<sup>9</sup> This perception was echoed in an interview with a couple from a small city in the Argentine province of Córdoba:

*Husband:* Here it's different than in Córdoba [the nearest big city]. Here they know everyone. And they know whom everyone is going to vote for.

*Author:* When people come and give things out during the campaign, are they people whom you know?

*Husband:* Yes, they're people from here, they're neighbors. Here everyone knows each other. "Small town, big hell." (*Pueblo chico, infierno grande.*)

*Author:* Do they know how you voted?

*Husband:* For many years we've seen, people will say, "So-and-so voted for so-and-so." And he wins, and they come and say, "You voted for so-and-so." I don't know how they do it, but they know.

*Wife:* We were at the *unidad básica* [a neighborhood Peronist locale] and they say to me, "[Your cousin] voted for Eloy" [the given name of a Radical-party candidate]. And I asked my cousin, "did you vote for Eloy?" And she said "yes"! They knew that my cousin had voted for Eloy!<sup>10</sup>

Voting technologies in Argentina also reduce the anonymity of the vote. Argentina has the secret but not the Australian ballot.<sup>11</sup> Argentines vote with slips of paper that carry the names only of a given party's candidates, like the coupon ballots used in the nineteenth-century United States. People can vote with ballots that they receive directly from party operatives. Or they can vote with ballots supplied inside the voting booth. People tend to receive ballots as part of a process of direct, face-to-face mobilization.

The practice of handing out ballots basically serves as a method of monitoring and influencing how people vote. One Peronist organizer explained in an interview

how the party used the ballots. "The most important thing is to go look for people and give them the ballot. You give them the ballot in the taxi [which the party has hired to transport them to the polls]. Then no one has time to change their ballots for them [i.e., give them a different ballot. After taking voters into the polling place] you put them on line to vote... Then they don't have a chance to change the ballot. Only if they're really sneaky and they change it inside the voting booth."<sup>12</sup>

In sum, my first assumption is that machines can effectively, if imperfectly, monitor the actions of their constituents.

A second assumption needed to model machine politics as a repeated game is that all players foresee the game continuing into the future. It is entirely appropriate to think of the interactions between machine operatives and their constituents as repeated over many iterations; the more artificial assumption would be that these are one-shot or short-lived interactions. Machines and clientelist parties are effective to the extent that they insert themselves into the social networks of constituents. The grassroots party operative is a long-time neighbor of the people she tries to mobilize. In Latin America, clientelist parties of renown have been long-standing organizations, deeply enmeshed in working-class communities: Peru's Partido Aprista Peruano (APRA), founded in the 1920s, Mexico's Institutional Revolutionary Party (PRI), founded in the 1930s, Argentina's Peronists, founded in the in the 1940s.

The repeated-play assumption may be most appropriate in countries where parties are old, even if the democracies in which they compete are new. The three democracies just mentioned are new: Peru and Argentina redemocratized in 1980 and 1983, respectively, and Mexico democratized for the first time in 2000. Yet clientelist parties in all of them are old. The repeated-play assumption may be less appropriate in new democracies where the major political parties are also young and hence less enmeshed in social networks.

When parties that are not enmeshed in social networks try to buy votes with private inducements, voters greet their efforts with skepticism. In connection with research I conducted in Lima, Peru, I observed the reactions of people in a working-class neighborhood to a soup kitchen that a political party established in 1985, shortly before national elections (see Stokes, 1995). Soup kitchens were familiar in the neighborhood: Catholic activists and women's organizations ran some and the local mayor's office supported them. But when residents saw an outsider party set up a soup kitchen they predicted that it would disappear after election day. They were unmoved by the sponsor-party's implicit appeal for electoral support. And they were right: the soup kitchen did disappear right after the election.

In the Argentine case, furthermore, it is appropriate to assume that parties and voters see their interaction

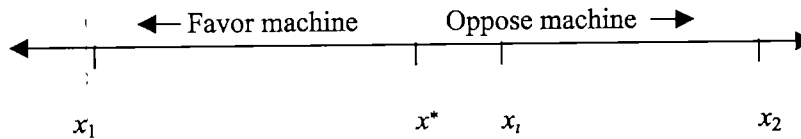
<sup>8</sup> Interview conducted January 2003, in the city of Córdoba, by Valeria Brusco, Marcelo Nazareno, and Susan Stokes

<sup>9</sup> We used multistage cluster sampling techniques, based on census tracks, to select 500 adults each in the provinces of Buenos Aires, Córdoba, Misiones, and San Luis. The margin of error was plus or minus 4.5%.

<sup>10</sup> Interview conducted by Valeria Brusco, Lucas Lázaro, and Susan Stokes, July 2003

<sup>11</sup> Scholars often fail to distinguish between the two. Argentina, Panama, and Uruguay are examples of developing democracies that don't use Australian ballots but where balloting is secret. Voting takes place in enclosed booths, and ballots are placed in opaque envelopes before being returned to election officials. But the ballots are produced by political parties and contain only a given party's list of candidates. Furthermore, I have cited two other developing democracies, India and Russia, where the Australian ballot is used but where experts claim that the secrecy of the ballot is informally violated.

<sup>12</sup> Interview conducted in June 2002 in the city of Córdoba, by Valeria Brusco, Marcelo Nazareno, and Susan Stokes

**FIGURE 1. The Location on a Spatial Dimension of a Political Machine ( $x_1$ ), Its Opponent ( $x_2$ ), the Median Voter ( $x^*$ ), and a Hypothetical Voter ( $x_i$ )**

as extending into the indefinite future; even if they could imagine hypothetical circumstances in which it might end (in the event, e.g., of a military coup), at the time of any given election since the return to democracy in that country, few would have anticipated a particular moment when it would end. The perception of an interaction with no identifiable stopping point makes it reasonable to model this as an infinitely repeated game.

To capture the repeated-play dynamic of machine politics, it is necessary to depart in a third way from received models of redistributive politics. These models assume that the machine's ability to reward voters for their support depends on its winning elections.<sup>13</sup> A voter whose support will only be rewarded if the machine wins anticipates that the game in effect ends each time the machine loses. Many machines, such as Mexico's PRI (Diaz-Cayeros, Magaloni, and Weingast 2001), Singapore's People's Action Party (PAP; Tam 2003), or, for many decades, Italy's Christian Democratic Party as it operated in the south (Chubb 1982), face negligible competition. Because the machine effectively cannot lose, voters anticipate that the game will continue. But other machines operate in settings where they can lose. Even in competitive settings, the game between machine and voter need not end when the machine finds itself in opposition. It does not end if the machine can carry over public funds from the party's time in power, or if it can make use of resources donated by private actors, private actors who expect policy concessions from the machine when it is back in power (Stigler 1975). Note that two of the three long-term clientelist Latin American parties mentioned earlier, the Peronists and APRA, were more often in opposition than in power.

To summarize, my key assumptions are that machines can monitor voters' actions and that both sides foresee their interaction extending indefinitely into the future. The latter assumption implies that machines don't lose their ability to distribute goods when they find themselves in opposition.

### The Model

I begin with a one-shot game in which a person's vote is assumed to be perfectly observable by political parties. Let the ideological position of the machine in a one-dimensional policy space be represented by  $x_1$ , the ideological position of the opposition by  $x_2$ , and  $x_1 < x_2$ . Let  $x^* \equiv (x_1 + x_2)/2$  be the midpoint between the two parties (see Figure 1). Let the voters' preferences be given by

$$u_i = -\frac{1}{2}(v_i - x_i)^2 + b_i,$$

where  $v_i = \{x_1, x_2\}$  represents a vote for either the machine or the opposition,  $x_i$  represents voter  $i$ 's position on the ideological spectrum, and  $b_i = \{0, b\}$  represents the value to the voter of the reward offered by the machine in exchange for votes, relative to the value of voting according to the voter's preferences. Thus  $-(1/2)(v_i - x_i)^2$  represents the expressive value of voting for one of the two parties. If the machine does not offer a gift, then  $b_i = 0$  and the voter votes for the machine if  $-(x_i - x_1)^2 \geq -(x_i - x_2)^2$ , or if  $x_i \geq x^*$ . That is, if there is no gift the voter supports the party that falls closest to the voter on the ideological or programmatic dimension. If the machine offers a gift of  $b > 0$ , the voter will vote for it if

$$-1/2(x_i - x_1)^2 + b \geq -1/2(x_i - x_2)^2,$$

or

$$b \geq \frac{1}{2}[(x_i - x_1)^2 - (x_i - x_2)^2] = (x_2 - x_1)(x_i - x^*),$$

or

$$x_i \leq x^* + (b/(x_2 - x_1)).$$

The normal form of the stage game is depicted in Table 1. In the Table, the machine is represented as expending  $b$  when it pays a reward, and gaining  $v$  when it receives a vote.

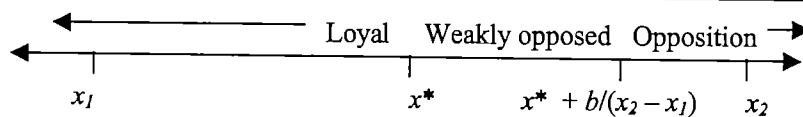
Define voters for whom  $x < x^*$  as *Loyal* voters (see Figure 2). Loyal voters' dominant strategy is to vote for the machine. Define voters for whom  $x > x^*$

<sup>13</sup> In static models of clientelism in which the party only pays a reward if it wins, a voter's actions depend on his or her beliefs about the likely actions of other voters. A collective-action problem arises when voters prefer, on programmatic grounds, to vote against the machine. Then defeating the machine is a public good, but individual voters pay a cost for attempting to unseat it if the attempt fails. See Medina and Stokes 2003, and Diaz-Cayeros, Magaloni, and Weingast 2001.

**TABLE 1. Normal Form of a Game Between the Machine Operative and a Voter**

Voter	Machine	
	Reward	No Reward
Comply	$-1/2(x_i - x_1)^2 + b, v - b$	$-1/2(x_i - x_1)^2, v$
Defect	$-1/2(x_i - x_2)^2 + b, -b$	$-1/2(x_i - x_2)^2, 0$



**FIGURE 2. Types of Voters by Their Location on a One-Dimensional Policy Space****TABLE 2. Normal Form of the Game Between the Machine Operative and the Weakly Opposed Voter with Simplified Payoffs**

Voter	Machine	
	Reward	No Reward
Comply	3, 3	1, 4
Defect	4, 1	2, 2

$b/(x_2 - x_1)$  as *Opposition* voters. Opposition voters will oppose the machine even if offered  $b$  to change their votes. Define voters for whom  $x^* < x < x^* + b/(x_2 - x_1)$  as *Weakly opposed* voters. Weakly opposed voters prefer to vote against the machine in the absence of a reward, but prefer to vote for the machine if doing so brings them a reward. If the value of the vote to the machine exceeds  $b$ , the machine and the Weakly opposed voter are in a prisoners' dilemma. Table 2 gives the game between a Weakly opposed voter and a machine, with simplified payoffs that make clear the prisoners'-dilemma structure of the game.

Next, I assume an infinite sequence of elections and model the interaction between the machine and a Weakly opposed voter as an iterated prisoners' dilemma with one-sided uncertainty.<sup>14</sup> I also assume that the two are playing a grim-trigger strategy, whereby when one player defects, the other defects in all subsequent rounds. Aside from theoretical reasons in favor of the grim trigger, interviews with Argentine party operatives suggest that they in fact follow a strategy of this sort. For instance, we asked a Peronist organizer how she responded when she suspected that a person to whom she had extended favors voted for another party. She answered, "He's dead. He died, forever."<sup>15</sup>

Returning to the model, if the voter votes against the machine, I now assume, the machine observes the negative vote with a probability  $p$ . Voters discount the future by a discount factor  $\beta$ , which falls on the interval  $[0, 1]$ . The condition for a subgame-perfect equilibrium (SPE) in which the Weakly opposed voter receives the

reward and votes for the machine, supported by a grim trigger strategy should the voter be observed to renege, is

$$\begin{aligned} & 1/(1 - \beta)[b - (x_i - x_1)^2/2] \\ & \geq [b - (x_i - x_2)^2/2] + [\beta/(1 - \beta)]\{(1 - p) \\ & \times [b - (x_i - x_1)^2/2] - p(x_i - x_2)^2/2\}. \end{aligned} \quad (1)$$

In other words, to sustain cooperation, the value to the voter in the current and all subsequent periods of voting for the machine and receiving a reward must equal or exceed the sum of the payoff from defecting in the current period plus (1) avoiding detection and returning to cooperation in the next and subsequent periods (with probability  $p$ ), or (2) being caught and, in all subsequent periods, voting against the machine but foregoing rewards (with probability  $1 - p$ ).

Inequality [1] simplifies to

$$x_i \leq x^* + \lambda(b/x_2 - x_1),$$

where

$$\lambda = p\beta/(1 - \beta + p\beta).$$

Hence, the set of voters who would sell their votes in exchange for a private benefit is the set whose ideal point,  $x_i$ , satisfies

$$x^* \leq x_i \leq x^* + \lambda(b/x_2 - x_1). \quad (2)$$

Lambda falls on the  $[0, 1]$  interval. Lambda is an increasing function of the discount rate ( $\beta$ ) and of the probability of a defector being caught ( $p$ ). If  $p = 0$  (there is no possibility that the machine would observe a defection by the voter), or if  $\beta = 0$  (the voter cares nothing about future consumption), then inequality [2] reduces to  $x_i = x^*$ . In these cases the machine can buy the votes only of voters who are indifferent, on ideological grounds, between the parties.

Loyal voters do not meet the condition in [2]. As illustrated in Figure 2, for Loyal voters  $x_L < x^*$ . Intuitively, Loyal voters who want to extract private rewards from their preferred party would, under the grim trigger, have to threaten to vote against the party forever if the machine denied them a reward once. Such a threat would lack credibility: the party knows that the Loyal voter, even without rewards, is better off cooperating forever than defecting forever.<sup>16</sup> Nor

<sup>14</sup> In a sense there is uncertainty on both sides, about whether the other will cooperate or defect in the future. This uncertainty characterizes all iterated prisoners' dilemmas—indeed, all repeated games—in which there is more than one equilibrium. I model this game as one of one-sided uncertainty because only the machine is uncertain about whether the voter has cooperated or defected. The voter, by contrast, observes perfectly whether the machine gives him a reward.

<sup>15</sup> Interview conducted in January 2003 in the city of Córdoba by Valeria Brusco, Marcelo Nazareno, and Susan Stokes

<sup>16</sup> The loyal voter's diehard ideological commitment to the party allows the machine, in a sense, to exploit him, garnering his vote without having to spend scarce resources on him. Loyalists would

do Opposition voters, those who oppose the machine on programmatic grounds more strongly than do the Weakly opposed, satisfy condition [2] (for Opposition voters,  $x_0 > x^* + \lambda(b/x_2 - x_1)$ ). The reason is that even though the Opposition voter would like to receive a reward, the machine cannot use the threat of withholding a reward to secure this voter's compliance: he is always better off forgoing the reward and voting against the machine. The machine knows this and does not offer him a reward.

Weakly opposed voters (and indifferent voters, where  $x_i = x^*$ ) are the only types whose policy ideal points make them potential vote sellers.<sup>17</sup> The intuition behind this result is that, in contrast to the Opposition voter, Weakly opposed voters can credibly commit to voting for the machine in exchange for a gift; the machine knows that the voter is better off cooperating forever than defecting forever. In contrast to the Loyal voter, the threat to punish the machine by voting against it in the future by the Weakly opposed voters is credible: left to their own devices, this is their preferred course of action.

Inequality [2] implies four comparative statics:

- As the ideological distance between the two parties ( $x_2 - x_1$ ) shrinks, the potential for vote buying grows. Intuitively, when the two parties are ideologically or programmatically close, there is less at stake for the voter in the decision of which to vote for, and the value of the private reward becomes more salient.
- As the value of the private reward ( $b$ ) relative to the value of voting in accordance to one's policy or ideological preference increases, the potential for vote buying increases. The reward must be worth a lot to the voter. But its value to the machine must be less than the value of a single vote—not very much. This suggests that, given decreasing marginal utility from income, machines will target poor voters.
- The more accurately the machine can monitor voters, the greater the potential for vote buying ( $\lambda$  is an increasing function of  $p$ ). This accuracy is a function of the technology for monitoring voters' actions and of the machine's organizational structure.
- Among its core constituents—those whom it can observe well—the machine is most effective when it targets Weakly opposed voters (for whom  $x^* \leq x_i \leq x^* + \lambda(b/x_2 - x_1)$ ), rather than Loyal ( $x_i < x^*$ ) or Opposition voters ( $x_i > x^* + b/(x_2 - x_1)$ ) voters.

therefore have an incentive to masquerade as indifferent voters, a possibility that I do not model here. It might, however, be psychologically difficult for party enthusiasts to feign indifference. Note also that any ideological shift by the machine runs the risk of turning the loyalist into an indifferent or even an opposition voter. Machines would then have to consider the distribution of loyal voters and the additional resources that might be needed to retain their support, were it to consider a change in its ideological stance.

<sup>17</sup> Their minmax payoffs are, for the machine, 0, and, for WO,  $-1/2(x_{WO} - x_2)^2$ . Hence, the feasible and individually rational payoffs they will accept in repeated play include the cooperation payoffs of  $(v - b, -1/2(x_{WO} - x_1)^2 + b)$

## MACHINE POLITICS AND VOTE BUYING IN ARGENTINA

The comparative statics from my formal model generate hypotheses about the causes of machine or clientelist politics. In this section, I test these hypotheses with evidence from one developing democracy, Argentina.<sup>18</sup> The evidence I present comes mainly from a survey of 1,920 voters, conducted in December 2001 and January 2002 in three Argentine provinces.<sup>19</sup> The survey allows us to explore the strategies of clientelist parties indirectly, by revealing what kinds of voters these parties target and who among the voters are responsive to private rewards.<sup>20</sup> Respondents were asked whether they had received any goods from a political party during the election campaign that had taken place two months earlier (variable name, *Reward*). Of low-income respondents in the sample, 12% (89 out of 734) reported having received goods. Most of them said that they had received food; other items mentioned frequently were building materials, mattresses, and clothing. In an open-ended question about whether receiving goods influenced their vote (*Influence*), about one in five of the low-income voters, and one-quarter of low-income Peronist voters, said it did. We asked other questions meant to detect clientelism, such as whether the person had turned to a locally important political actor for help during the past year (*Patron*) and whether, if the head of their household lost his or her job, the family would turn to a party operative for help (*Job*).

### Poverty and Vote Buying

I discuss five pieces of evidence from the survey that lend support to my theory of machine politics. The first has to do with the effect of poverty on a voter's willingness to sell his or her vote. The formal model analyzed earlier predicts that vote buying is more easily

<sup>18</sup> The one comparative static from the model that I do not test is that ideological proximity between the parties encourages vote buying. The surveys did not elicit respondents' views of the ideological distance between Argentina's two major parties.

<sup>19</sup> As in the 2003 survey reported on earlier, we used multistage cluster sampling techniques, based on census tracks. In this earlier survey we selected 480 adults each in the provinces of Buenos Aires, Córdoba, and Misiones, and from the area of Mar del Plata. The margin of error was plus or minus 4.5%.

<sup>20</sup> Students of political clientelism and redistributive politics have typically observed the distribution of resources and their effects on voting at aggregated levels, such as the district or the county (see, e.g., Ansolabehere and Snyder, 2002, or Diaz-Cayeros, Magaloni, and Weingast, 2001). The problem of ecological inference can mar this approach. In contrast, the main problem with the survey approach used here is that people may be reluctant to acknowledge receiving handouts, in the Argentine case probably as much because of the implication that they are poor enough to sell their votes as out of concern about the illegality or immorality of their actions. It is probably evidence of this reluctance that only 7% of our sample acknowledged having received goods, whereas 44% said goods were distributed in their neighborhood, 39% could mention exactly what items were distributed, and 35% could name the party that gave them out. The effect of underreporting of clientelism is, in estimations where it is the dependent variable, to bias coefficients downward and make statistically significant associations appear insignificant.

sustained, all else equal, when the voter values the private reward relatively highly but the party values it relatively little. The picture this paints is of parties giving minor benefits to voters who are poor enough to value them highly—a picture consistent with much of the qualitative literature on machine and clientelist parties. To cite just one of many examples, Wilson and Banfield (1963) explain that U.S. machines operated in a city's "river wards," where working-class residents lived, but not in the "newspaper wards," where middle-class residents lived.

Table 3 reports regression estimates of the likelihood of a clientelistic response to the set of questions discussed earlier, including whether the respondent received a private reward from a party. The negative and significant coefficients on *Income*, *Education*, and *Housing quality* variables show that poverty predicts clientelism. To illustrate the effect, the simulated expected probability that a wealthy person (one with the highest income, education, and housing-quality level) would have received a reward and acknowledged that it influenced her vote is 0.2%. The probability that a poor person (one with the lowest income, education, and housing-quality level) would have received a reward and allowed his or her vote to be influenced by it is 65 times greater: 13%.<sup>21</sup>

In sum, political machines buy the votes of poor people in Argentina.

## Monitoring Voters

**Machine Organizational Structure.** In the presence of the secret ballot, parties make inferences about how people vote by observing their type—where they fall on the dimension of programmatic support for the parties. A tentacle-like organizational structure is a great asset to parties in this regard. We know from a large secondary literature that the Argentine party with the organizational structure most like that of the machine is the Peronist party (see, e.g., Auyero 2000, and Levitsky 2003). And our surveys indicate that the Peronist party was by far the most active in distributing private rewards. Eight hundred thirty-nine of our respondents said that a party distributed private rewards in their neighborhoods during the campaign; of these, 423 (50%) said that the Peronists distributed them. The next most frequently mentioned party, the Radical Party, was mentioned by only 49 respondents.

<sup>21</sup> All simulations reported in this section were executed with the *Clarify* program (King, Tomz, and Wittenberg, 2000, and Tomz, Wittenberg, and King, 2001). *Clarify* draws simulations of parameters of statistical models (in this case, ordered logit regressions) from their sampling distribution and then converts these simulated parameters into expected values, such as expected probabilities of an answer to a survey question, given hypothetical values of explanatory variables. *Clarify* software and documentation are available from Gary King's web site at <http://gking.harvard.edu>. For this simulation I assumed a female Peronist supporter whose age and municipality size were average for the sample. Confidence intervals around the 0.2% expected probability were 0.05% and 0.5%, and around the 13% probability, 7% and 22%.

**TABLE 3. Model Estimations of Vote Buying**

	(1)	(2)	(3)	(4)
Dependent Variable	Patron	Job	Reward	Influence
Model	Logit	Logit	Logit	Ordered Logit
Estimated				
Income	<b>-0.126</b> (0.058)	-0.054 (0.037)	<b>-0.195</b> (0.074)	<b>-0.194</b> (0.070)
Education	-0.005 (0.058)	<b>-0.197</b> (0.035)	<b>-0.212</b> (0.079)	<b>-0.223</b> (0.073)
Housing quality	-0.215 (0.114)	<b>-0.133</b> (0.073)	-0.212 (0.131)	<b>-0.310</b> (0.022)
Log population	<b>-0.361</b> (0.044)	-0.035 (0.029)	<b>-0.135</b> (0.050)	<b>-0.139</b> (0.045)
Ballot			<b>0.578</b> (0.225)	<b>0.572</b> (0.211)
Peronist sympathizer	<b>0.594</b> (0.192)	<b>0.735</b> (0.119)	<b>0.550</b> (0.220)	<b>0.549</b> (0.207)
Age	-0.005 (0.006)	<b>-0.022</b> (0.003)	<b>-0.016</b> (0.007)	<b>-0.017</b> (0.006)
Gender	-0.178 (0.166)	<b>0.208</b> (0.103)	-0.158 (0.195)	0.092 (0.180)
Radical sympathizer	<b>0.357</b> (0.243)	0.146 (0.158)	-0.455 (0.371)	0.026 (0.299)
Constant	<b>3.254</b> (0.643)	<b>1.879</b> (0.397)	<b>1.580</b> (0.746)	
N observations	1114	1920	1618	1619

Note. Cell entries are coefficients, and standard errors are in parentheses. Boldface indicates significance at the  $p=0.05$  level or smaller.

Explanation of dependent variables: *Patron*: "In the past year, have you turned to [the person the respondent previously identified as the most important local political figure] for help?" Coded yes = 1. *Job*: "If the head of your household lost his or her job, would you turn to a party operative for help?" Coded yes = 1. *Reward*: "Did you receive goods distributed by a party in the last campaign?" Coded yes = 1. *Influence*: "Did the fact of having received goods influence your vote?" Coded 1 = Did not receive goods, 2 = received goods, no influence; 3 = received goods, acknowledged influence. Based on responses to open-ended question.

Explanation of independent variables: *Log population*: natural log of population of respondent's municipality (2001 census). *Ballot*: coded 1 for people who reported voting with a ballot given to them by a party operative, 0 for people who voted with a ballot they acquired in the voting booth. *Peronist sympathizer*: coded 1 for respondents who said they liked the Peronist Party more than others, 0 otherwise. *Income*: Self-reported by respondent, 9-level scale. *Education*: 9-level scale, from no formal education to postgraduate. *Housing quality*: Assessed by interviewer, 5-level scale (1 = poorest quality, 5 = highest quality). *Gender*: female = 1. *Radical sympathizer*: coded 1 for respondents who said they liked the Radical Party more than others, 0 otherwise.

**Community Structure.** The ease of monitoring is also influenced by the structure of communities where machines operate. We expect voters to be less anonymous, their partisan predispositions or types more a matter of public knowledge, in smaller towns and cities, where social relations are multifaceted and where, as one person we interviewed put it, "everyone knows each other."<sup>22</sup> These are places where it is easier for parties to know

<sup>22</sup> Interview conducted by Valeria Brusco, Lucas Lázaro, and Susan Stokes, July 2003.

who's who, who is inclined toward one party or another, and how people are likely to have voted. And they are places where, as the same person explained, parties can use this information to "discriminate a little" when a defecting voter comes and "asks for a favor." It is reasonable, then, to treat community size as a proxy for observability of residents' votes.

In our surveys, the smaller the population size of the respondent's municipality, the more likely she or he was to have received rewards and to be responsive to them. These two effects are revealed in Table 3 by the negative and significant coefficients relating logged population size (as measured in the 2001 census) to *Patron Reward* and to *Influence* variables.<sup>23</sup>

**The Technology of Voting.** The last two findings—that rewards are distributed by the party with the most machine-like structure, and that people in small towns and cities are more likely to receive, and to be responsive to, rewards—might be interpreted as simply showing that parties hand out rewards preferentially to people whom they can reach most efficiently. But I have argued that efficiency of distribution is just one side of the link between political machines and their constituents. The other side is perverse accountability: the machine's ability to hold voters accountable for their votes.

The fourth piece of evidence that I report goes directly to a party's ability to discern people's votes and to condition rewards on compliance. This evidence has to do with the technology of voting. Recall that Argentines vote with party-produced ballots, which they can acquire either directly from party operatives, as part of the process of face-to-face mobilization, or anonymously, in the voting booth. *Ballot* in Table 3 is a dummy variable for people who voted with ballots given to them by party operatives (15% of our sample). The positive and significant coefficient relating *Ballot* to *Reward* shows that people who vote with personally distributed ballots are more likely than others to receive rewards from parties, such as food or clothing. The positive and significant coefficient relating *Ballot* to *Influence* shows that people who receive personally distributed ballots are also more responsive to rewards.<sup>24</sup>

To give a sense of the magnitude of this effect, the simulated expected probability that a poor voter would allow his or her vote to be influenced by a reward, as we

have seen, is 13%. This assumes that the voter received his or her ballot from a party operative. If we assume the same hypothetical poor voter voted with a ballot he or she finds in the voting booth, the probability is cut almost in half, to 7%.

In sum, in Argentina the more able a party is to monitor its constituents, the more effective its efforts at vote buying. The party with the most decentralized and tentacle-like organizational structure, hence the one best able to monitor the actions and types of its constituents, was the party that most actively attempted to buy votes. And the more observable the vote, either because the voter lives in a small community or because he or she receives the ballot directly from a party operative, the more likely he or she is to be the target of vote buying.

## Types of Voters and Vote Buying

A fifth piece of evidence speaks to the question, What types of voters do machines pursue? My theoretical prediction was that machines focus their vote-buying efforts on people in the middle of the distribution of partisan predispositions: ones who are indifferent about whether to vote for or against the machine ( $x_M = x^*$ ), and ones with a weak predisposition against it ( $x^* \leq x \leq x^* + b/(x_2 - x_1)$ ). Machines will avoid voters who are loyalists or strong opponents.

We asked respondents their opinions of the Peronist party, Argentina's preeminent political machine. We asked them to choose among "very good," "good," "bad," and "very bad" as their answers. Figure 3 displays the percentages of people who received or did not receive handouts by opinions of the Peronist party.

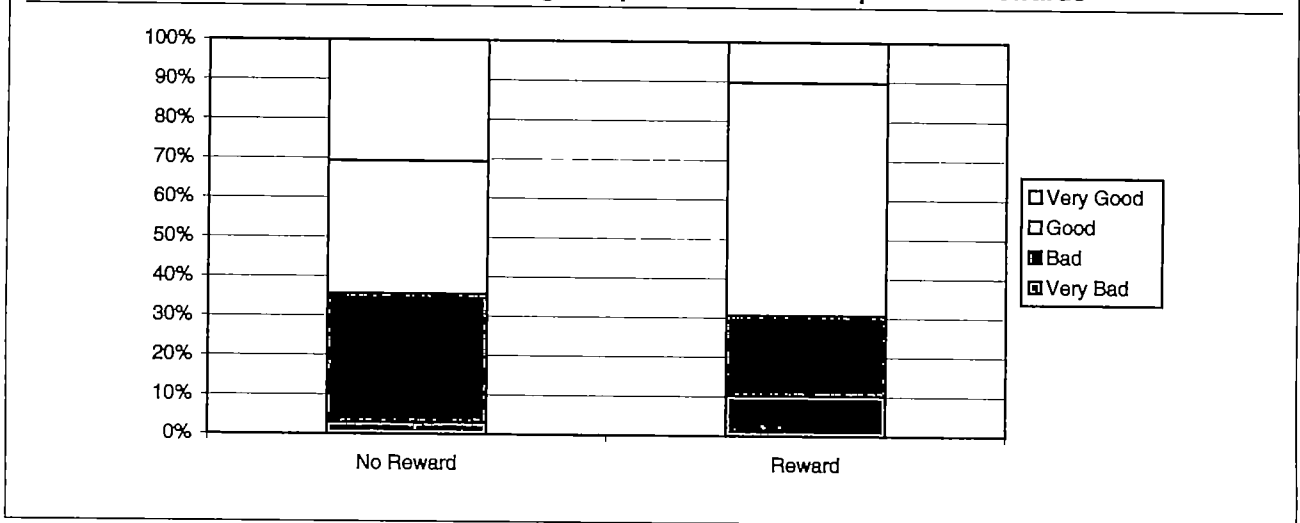
A striking finding, and one that conforms to the theoretical prediction, is the small proportion of those who rated the party "very good" who received rewards. Three times as many people who did *not* receive rewards as those who *did* receive them rated the Peronists "very good" (31% vs. 10%). The Peronist party turned away from its strongest loyalists when it gave out private rewards. (The difference is all the more striking given that one might anticipate some endogeneity of perceptions of the party: people who receive rewards from it might be more prone, because of the gift, to rate the party "very good.") Another aspect of the findings that accords with my model's predictions is that many more people who rated the party "bad" received rewards than those who rated it "very bad."

In some ways, however, the findings do not accord with the predictions. Recipients of rewards were concentrated in the "Good" category: nearly 60% of those who received handouts from the Peronists saw it as a good party. These findings are inconsistent with the theory if we think of people who called the machine "good" as falling somewhat to the left of the median ideal point ( $x^*$ ) in Figure 2 and hence as being weakly predisposed in the machine's favor. Recall that, in theory, even voters just mildly predisposed in the machine's favor would not be able to credibly threaten to punish the machine if it defected and therefore would not, in repeated play, be able to induce the machine

<sup>23</sup> Note that 90% of our interviews were with people who lived in cities with more than 10 thousand inhabitants. Thus, we interviewed few people who could be said to live in rural communities, and our population variable is best interpreted as distinguishing people according to the size of the urban area in which they lived.

<sup>24</sup> The confidence intervals around the 7% figure are 4% and 12%. An alternative interpretation is that parties, as a service, deliver ballots to the loyal partisans, who are more likely to vote for them anyway. In this case partisanship would "cause" both the hand delivery of the ballot and support for the party, and the apparent link between ballot delivery and support would be spurious. Yet this alternative explanation is inconsistent with the testimony of party operatives, who, like the one cited earlier, focus their ballot-delivery efforts on uncommitted or indifferent voters, ones who—they fear—might change the ballot in the voting booth.



**FIGURE 3. Opinions of Peronists Among Recipients and Nonrecipients of Rewards**

to pay them rewards. And we might have expected a relatively larger proportion of voters who rated the Peronists as “bad,” and hence who were weakly opposed to it, to receive rewards. Similarly, note that the regression models in Table 3 show that, controlling for other factors, Peronist sympathizers were significantly more likely to receive rewards.

One explanation for the slippage between the theory and the evidence is that our survey did not offer people the option of indicating true indifference about the Peronists. Some people who chose the “good” option might in fact be closer to indifferent. And some people who were close to indifferent, prerewards, might have called the party “bad” but, because of the reward, been nudged into seeing it as “good.”

The finding may also suggest a dynamic that goes beyond the model. Political machines organize by neighborhood and district, and they do more than just give out tactical rewards. They also proselytize. Although their proselytizing, in a competitive setting such as Argentina’s, is not perfectly successful, to the extent that it is successful at all we expect the distribution of voter types in areas of machine organizational penetration to be skewed toward machine supporters, weak and strong. In other words, we expect organizational penetration by the party to increase not only the efficiency with which it distributes rewards and its ability to monitor voters, but also its partisan support (as Cox and McCubbins 1986 assume). If organizational penetration increases partisan support, then the machine will target its supporters more than its opponents simply because it has greater access to them. Whatever the explanation for this anomaly, the evidence from Argentina does show unambiguously that, among core constituents, the machine discriminates against its most ardent supporters.

## CONCLUSIONS

The dynamic model I analyze and test here by no means answers all of our questions about machine

politics. For reasons of space, I haven’t addressed the question, If two parties compete by offering private rewards, what determines a voter’s choice? One can imagine a bidding-war dynamic, where the value of private rewards escalates rapidly. If two parties offered private rewards of the same value, one would expect the machines to compete for the same set of (ideologically) marginal voters. But competition between “dueling machines” seems, empirically, unusual. It is more common that, even in settings where politics is competitive at the macro level, parties have especially close links to particular groups of voters. And often one party specializes in machine-style politics, whereas another focuses on programmatic mobilization.

This last point raises the question, If parties that are organized as machines can use minor payoffs to sway voters, why don’t all parties organize themselves this way? A tentative answer is that parties face unequal costs of monitoring voters. Monitors are most effective when they live among the voters they are observing. Given residential segregation by income, parties with a middle-class base would have to employ middle-class monitors, who would require greater compensation than do the working-class operatives. Parties with middle-class constituencies therefore are more effective when they advertise their programs, focusing resources on “air,” rather than “ground,” campaigns.

These limitations notwithstanding, we have made some headway. I have returned to Scott’s insight that machine politics of old is a lot like clientelist politics of new. I have argued that the dynamics of machine or clientelist redistribution has only been half-understood in the literature, which has captured the delivery-of-services but not the monitoring-of-voters side of the story. The literature thus misses the fact that machines are able to use their social proximity to voters to monitor their actions and types and hence to enforce the implicit redistributive contract. This insight allows us to model the strategic interactions between machines and constituents as repeated games.

Casting voter-machine interactions as repeated games allows us to overcome commitment problems, over which the formal literature has stumbled, and to identify equilibria in which vote buying actually takes place. I have shown formally that when voters see parties as ideologically close to one another, vote buying is more likely to occur. I have shown, formally and empirically, that machines target poor people, for whom the payoff of even a small reward outweighs the expressive value of voting for one's preferred party. Empirical evidence also supports the theoretical finding that the more accurately the machine monitors individual voters, either through a tentacle-like party structure or through voting technologies that reduce the anonymity of the vote, the more successful are its efforts at vote buying. And evidence supports (though with some nuances) the theoretical finding that machines avoid extending largesse to diehard loyalists and focus their rewards on voters in the middle of the distribution of partisanship.

The Argentine evidence, then, on the whole supports the theoretical finding that perverse accountability—the ability of parties to monitor constituents' votes, reward them for their support and punish them for defection—is what sustains machine politics.

## REFERENCES

- Ansolahehere, Stephen, and James M. Snyder, Jr. 2002. Party Control of State Government and the Distribution of Public Expenditures. Massachusetts Institute of Technology. Typescript.
- Auyero, Javier. 2000. *Poor People's Politics: Peronist Survival Networks and the Legacy of Evita*. Durham, NC: Duke University Press.
- Calvo, Ernesto, and Maria Victoria Murillo. 2004. "Who Delivers? Partisan Clients in the Argentine Electoral Market." *American Journal of Political Science* (October): 742–57.
- Chandra, Kanchan. 2004. *Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India*. New York: Cambridge University Press.
- Chubb, Judith. 1982. *Patronage, Power, and Poverty in Southern Italy*. New York: Cambridge University Press.
- Cox, Gary W., and Matthew D. McCubbins. 1986. "Electoral Politics as a Redistributive Game." *Journal of Politics* 48 (May): 370–89.
- Diaz-Cayeros, Alberto, Beatriz Magaloni, and Barry Weingast. 2001. *Democratization and the Economy in Mexico: Equilibrium (PRI) Hegemony and its Demise*. Stanford University. Typescript.
- Dixit, Avinash, and John Londregan. 1996. "The Determinants of Success of Special Interests in Redistributive Politics." *Journal of Politics* 58 (November): 1132–55.
- Fox, Jonathan. 1994. "The Difficult Transition from Clientelism to Citizenship: Lessons from Mexico." *World Politics* 46 (January): 151–84.
- Karlan, Pamela. 1994. "Not by Virtue but by Money Won? Vote Trafficking and the Voting Rights Systems." *Virginia Law Review* 80 (October): 1455–75.
- Keyssar, Alexander. 2000. *The Right to Vote: The Contested History of Democracy in the United States*. New York: Basic Books.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analysis: Improving Interpretation and Presentation." *American Journal of Political Science* 94: 347–61.
- Kitschelt, Herbert, Zdenka Mansfeldova, Radoslaw Markowski, and Gábor Tóka. 1999. *Post-Communist Party Systems: Competition, Representation, and Inter-Party Cooperation*. Cambridge: Cambridge University Press.
- Kochin, Michael S., and Levis A. Kochin. 1998. "When Is Vote Buying Wrong?" *Public Choice* 97: 645–62.
- Levitsky, Steven. 2003. *Transforming Labor-Based Parties in Latin America: Argentine Peronism in Comparative Perspective*. New York: Cambridge University Press.
- McCaffery, Peter. 1993. *When Bosses Ruled Philadelphia: The Emergence of the Republican Machine, 1867–1933*. University Park, PA: Penn State Press.
- Medina, Luis Fernando, and Susan C. Stokes. 2003. *Clientelism as Political Monopoly*. Chicago Center on Democracy Working Paper #25.
- O'Donnell, Guillermo. 1996. *Another Institutionalization Latin America and Elsewhere*. Kellogg Institute Working Paper #222. Notre Dame, IN: Kellogg Institute for International Studies.
- Organization for Security and Cooperation in Europe. 2003. *Russian Federation State Duma Elections, 7 December 2003. Preliminary Findings of the OSCE International Election Observation Mission*. Vienna, Austria: OSCE.
- Piattoni, Simona, ed. 2001. *Clientelism, Interests, and Democratic Representation. The European Experience in Historical Perspective*. Cambridge, England: Cambridge University Press.
- Reynolds, John. 1988. *Testing Democracy: Electoral Behavior and Progressive Reform in New Jersey, 1880–1920*. Chapel Hill, NC: University of North Carolina Press.
- Reynolds, John. 1980. "The Silent Dollar". Vote Buying in New Jersey." *New Jersey History* 98: 191–211.
- Scott, James C. 1969. "Corruption, Machine Politics, and Political Change." *American Political Science Review* 63 (December): 1142–58.
- Stigler, George J. 1975. *The Citizen and the State. Essays on Regulation*. Chicago: University of Chicago Press.
- Stokes, Susan C. 1995. *Cultures in Conflict. Social Movements and the State in Peru*. Berkeley, CA: University of California Press.
- Stokes, Susan C. 2004. *Is Vote Buying Undemocratic?* Paper presented annual meetings of the American Political Science Association, Chicago.
- Tam, Waikeng. 2003. *Clientelist Politics in Singapore: Selective Provision of Housing Services as an Electoral Mobilization Strategy*. University of Chicago. Typescript.
- Tomz, Michael, Jason Wittenberg, and Gary King. 2003. *Clarify: Software for Interpreting and Presenting Statistical Results. Version 2.1*. Cambridge, MA: Harvard University, January 5. <http://gking.harvard.edu>.
- Wilson, James Q., and Edward Banfield. 1963. *City Politics*. Cambridge, MA: Harvard University Press.

# A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective

TULIA G. FALLETI *University of Pennsylvania*

**B**oth advocates and critics of decentralization assume that decentralization invariably increases the power of subnational governments. However, a closer examination of the consequences of decentralization across countries reveals that the magnitude of such change can range from substantial to insignificant. In this article, I propose a sequential theory of decentralization that has three main characteristics: (1) it defines decentralization as a process, (2) it takes into account the territorial interests of bargaining actors, and (3) it incorporates policy feedback effects. I argue that the sequencing of different types of decentralization (fiscal, administrative, and political) is a key determinant of the evolution of intergovernmental balance of power. I measure this evolution in the four largest Latin American countries and apply the theory to the two extreme cases (Colombia and Argentina). I show that, contrary to commonly held opinion, decentralization does not necessarily increase the power of governors and mayors.

**D**oes decentralization always increase the power of governors and mayors? If so, what explains the different degrees of change observed in the intergovernmental balance of power? Over the last 30 years, decentralization reforms have swept across the world, changing decades of centralized political and economic practices as well as the way in which we study politics. As James Manor writes, "Nearly all countries worldwide are now experimenting with decentralization . . . seen as a solution to many different kinds of problems" (1999, vii). One need only look as far as the fiscal data to observe this trend. In 1980, subnational governments around the world collected on average 15% of revenues and spent 20% of expenditures. By the late 1990s, those figures had risen to 19% and 25%, respectively, and had even doubled in some regions.<sup>1</sup>

Tulia G. Falletti is Assistant Professor, Department of Political Science, University of Pennsylvania, 208 South 37th Street, 202 Stetler Hall, Philadelphia, PA, 19104-6215 (falletti@sas.upenn.edu).

The author thanks Isabella Alcañiz, Dawn Brancati, Ruth Ben-Artzi, Ernesto Calvo, Maxwell Cameron, Edward Gibson, Gerald McDermott, Richard O. Moore, Ben R. Schneider, Rogers Smith, Kathleen Thelen, and the editor of the *American Political Science Review* and three anonymous reviewers for very helpful comments on previous drafts. The author is also grateful to Daniel Brinks, Katrina Burgess, Barbara Geddes, John Gerring, Merilee Grindle, Wendy Hunter, José Itzigsohn, James Mahoney, Paul Pierson, Dietrich Rueschemeyer, Theda Skocpol, Benjamin Smith, Martin Tanaka, and the seminar participants at Northwestern University, Brown University, the Central European University in Budapest, the David Rockefeller Center for Latin American Studies at Harvard University, the Kellogg Institute at the University of Notre Dame, and the migration of authority conference at the University of Michigan-Ann Arbor for their comments at earlier stages of the project; and to David Samuels and Matthew Shugart for sharing their congressional data on Brazil and Colombia. The Social Science Research Council, the United States Institute of Peace, the Center for International and Comparative Studies and the Alumpae Association at Northwestern University supported the fieldwork on which this article is based. The Kellogg Institute for International Studies at the University of Notre Dame and the Centre of International Relations at the University of British Columbia provided financial support and outstanding environments to write this article.

<sup>1</sup> In the large Latin American countries (Argentina, Bolivia, Brazil, Chile, Colombia, Mexico, Paraguay, and Peru, for which comparable data is available) the subnational shares of revenues and expenditures increased from averages of 14% and 16% in 1980, respectively,

Moving beyond the fiscal arena, the decentralization movement has seen major public services such as education and health transferred to subnational governments. Moreover, political and electoral reforms have left governors and mayors more accountable to their constituencies. This large-scale transfer of resources, responsibilities, and authority has brought subnational governments to the forefront of politics. Recent international news headlines testify to the importance of subnational elections and local governance issues.<sup>2</sup> The decentralization movement has also highlighted the relevance of intergovernmental relations, once described as the "hidden" or "fourth branch of government" (Edmund Muskie 1962, cited in Wright 1978, 5), in comparative politics. Increasingly, political scientists are shifting the locus of their analyses from the national to the subnational levels (Snyder 2001) and from the horizontal relations among branches of government to the vertical relations between levels of government (Gibson 2004). Despite this ostensible change in the political and analytic landscapes, the question remains, has decentralization led to the expected shift in the balance of power among presidents, governors, and mayors?

A substantial body of work on the consequences of decentralization hinges on the answer to this question; nevertheless, little attention is paid in the literature to a critical assumption that could very well be unjustified. Political scientists who draw from the liberal tradition argue that decentralization helps to deepen and consolidate democracy by devolving power to local governments (Diamond and Tsalik 1999). Economists who draw from a market theory of local expenditures argue that decentralization helps to improve resource allocation through better knowledge of local preferences and competition among localities (Oates 1972). Other scholars, meanwhile, warn against the devolution of

to 29% in 2000. Source data available at: <http://www1.worldbank.org/publicsector/decentralization/fiscalindicators.htm>.

<sup>2</sup> "The waiters' revolt. State elections in Mexico," *The Economist*, February 12, 2005; "Conservatives Claim to Carry German State in Close Vote," *The New York Times*, February 21, 2005; among others.

power to subnational officials and show that it can augment distributional conflicts (Treisman 1999), foster subnational authoritarianism (Cornelius, Eisenstadt, and Hindley 1999), and exacerbate patronage (Samuels 2003). Recent studies also suggest that, in the absence of proper fiscal and political mechanisms, the transfer of resources to subnational governments may lead to higher levels of inflation (Treisman 2000), larger deficits (Rodden 2002), and poorer overall macroeconomic performance (Wibbels 2000). Interestingly, despite their disagreements on the effects of decentralization for democratization and economic reform, all of the aforementioned studies share an assumption that decentralization increases the power of subnational officials. This power increase is generally used as the intervening variable connecting decentralization policies and either positive or negative outcomes, without questioning the existence of such a power increase in the first place.

If we conceive of decentralization as a multidimensional process (Montero and Samuels 2004, 8) that entails political bargaining over the content and implementation of different types of policies, we find that certain forms of decentralization in fact *decrease* the power of subnational officials. In order to evaluate the consequences of decentralization on broader processes of democratization and economic reform, we need to establish first when and how decentralization policies increase or decrease the power of subnational officials. This article advances a definition of decentralization that distinguishes among administrative, fiscal, and political decentralization. Unlike previous studies that have, for the most part, treated these categories separately, the definition presented here allows a distinction to be made between decentralization processes that increase the power of subnational officials and those that—contrary to the expectation—do not.<sup>3</sup> Furthermore, because we lack a framework to understand how the transfer of authority in one area interacts with, reinforces, or halts decentralization reforms in other areas, this article studies the interactions among different types of decentralization as they evolve over time.

By drawing on recent works on path dependence and institutional change (e.g., Mahoney 2000; Pierson 1992, 2000, 2004; Thelen 2000, 2003), this article provides a dynamic analysis of decentralization. In this approach, the conditions under which decentralization is first implemented and the timing and order of the

policies are important determinants of the evolution of intergovernmental balance of power. Previous studies have successfully accounted for varying degrees of fiscal decentralization at one point in time (e.g., Garman, Haggard, and Willis 2001), but have fallen short of explaining the effects of decentralization policies on the evolution of intergovernmental relations. I will not only measure the absolute level of decentralization at different points in time but also trace the effects of earlier reforms on later ones.

The article also brings subnational actors and interests to the center of the analysis. The puzzle of why national politicians choose or agree to give power away has led scholars to focus largely on the interests of national politicians toward decentralization, either in the executive branch (Grindle 2000; O'Neill 2003) or in the relations between the national executive and the legislature (Escobar-Lemmon 2003; Willis, Garman, and Haggard 1999). I show that a wide array of social and political actors, including the governors and their ministers, the mayors, the governors' and mayors' associations, the unions of the sectors to be decentralized, and other sectors of civil society are also the makers of decentralization.

Finally, the article emphasizes the territorial component of interest representation. A large part of the literature on decentralization has focused on the partisan or electoral incentives that move decentralization forward. Although very important, such emphasis on electoral incentives overlooks the territorial aspects of interest representation. In issues of decentralization, the territorial interests that derive from the choice of officials through geographic areas (Tarrow 1978, 4) are as important as electoral incentives. I show later that the feasibility and contents of decentralization reforms do not lie solely with politicians' electoral calculations, but also with their territorial interests. Thus, types of decentralization, territorial interests, and sequences of reforms are the three main components of the sequential theory advanced in the following sections. This theory will serve to explain when and why decentralization policies are likely to either increase or decrease the power of subnational officials.

## SEQUENTIAL THEORY OF DECENTRALIZATION

### Decentralization as a Process

Decentralization is a process of state reform composed by a set of public policies that transfer responsibilities, resources, or authority from higher to lower levels of government in the context of a specific type of state. Compared to previous definitions, this one poses four important restrictions. First, decentralization is conceived as a process of public policy reforms and not as a description of the state of being of the political or fiscal systems at a point in time. Second, lower levels of government are the recipients of the transferred responsibilities, resources, or authority. Reforms such as privatization or deregulation, which target nonstate

<sup>3</sup> By prioritizing different theories and methodological approaches, the literature on decentralization has divided the process into its component parts. Policy-oriented works have undertaken the study of administrative reforms, such as the transfers of education and health services (e.g., Di Gropello and Cominetti 1998). Another group of works has sought to explain the reasons behind political decentralization or why rational actors choose to give power away (Grindle 2000; O'Neill 2003). Likewise, institutional approaches have argued that differences in the political party systems explain the degrees of fiscal or political decentralization (Riker 1964; Willis, Garman, and Haggard 1999). Few studies have analyzed two or three types of decentralization at the same time (e.g., Manor 1999, Penfold-Becerra 1999), but even these studies do not analyze the interactions among the different policies and the consequences of their timing and evolution.



actors, are not included in this definition (cf. Cheema and Rondinelli 1983, 24–5). Third, because decentralization is a process of state reform, a transition to a different type of state necessarily implies the commencement of a new decentralization sequence. The contents of decentralization policies and their interaction with the broader political and economic systems are highly determined by the type of state they seek to reform. Hence, in order to compare decentralization policies across countries as part of analytically equivalent processes, we must compare policies taking place within the *same* type of state. Finally, in studying the downward reallocation of authority, much is gained from a clear taxonomy of decentralization based on the type of authority devolved, such that three types of decentralization can be distinguished:<sup>4</sup>

- *Administrative decentralization* comprises the set of policies that transfer the administration and delivery of social services such as education, health, social welfare, or housing to subnational governments. Administrative decentralization may entail the devolution of decision-making authority over these policies, but this is not a necessary condition. If revenues are transferred from the center to meet the costs of the administration and delivery of social services, administrative decentralization is funded and coincides with a fiscal decentralization measure. If subnational governments bear these costs with their own pre-existing revenues, administrative decentralization is not funded.
- *Fiscal decentralization* refers to the set of policies designed to increase the revenues or fiscal autonomy of subnational governments. Fiscal decentralization policies can assume different institutional forms such as an increase of transfers from the central government, the creation of new subnational taxes, or the delegation of tax authority that was previously national.<sup>5</sup>
- *Political decentralization* is the set of constitutional amendments and electoral reforms designed to open new—or activate existing but dormant or ineffective—spaces for the representation of subnational polities. Political decentralization policies are designed to devolve political authority or electoral capacities to subnational actors. Examples of this type of reforms are the popular election of mayors and governors who in previous constitutional periods were appointed, the creation of subnational legislative assemblies, or constitutional reforms that

strengthen the political autonomy of subnational governments.<sup>6</sup>

Regarding the consequences of each type of decentralization, I expect administrative decentralization to have either a positive or a negative impact on the autonomy of subnational executives. If administrative decentralization improves local and state bureaucracies, fosters training of local officials, or facilitates learning through the practice of delivering new responsibilities, it will increase the organizational capacities of subnational governments. Nevertheless, if administrative decentralization takes place without the transfer of funds, this reform may decrease the autonomy of subnational officials, who will be more dependent on subsequent national fiscal transfers or subnational debt for the delivery of public social services. Similarly, fiscal decentralization can have either a positive or a negative impact on the degree of autonomy of the subnational level. The result will depend largely on the design of the fiscal decentralization policy implemented. Higher levels of automatic transfers increase the autonomy of subnational officials because they benefit from higher levels of resources without being responsible for the costs (political and bureaucratic) of collecting those revenues. On the contrary, the delegation of taxing authority to subnational units that lack the administrative capacity to collect new taxes can set serious constraints on the local budgets and increase the dependence of the local officials on the transfers from the center. Prosperous subnational units prefer to collect their own taxes, but poor states or municipalities are negatively affected every time the collection of taxes is decentralized and, as a consequence, the horizontal redistribution of transfers from rich to poor subnational units is affected. Finally, political decentralization, by the definition provided previously, should almost invariably increase the degree of autonomy of subnational officials from the center. The only case when political decentralization could have a negative effect on the power of governors and mayors vis-à-vis higher level authorities is when, by augmenting the separation of powers at the subnational level (such as through the creation of subnational legislatures or municipal councils), it leads to divided subnational governments. In such instances, the subnational political opposition could undermine the authority of governors and mayors vis-à-vis the national executive.

<sup>4</sup> I do not distinguish among policies according to the degree of authority devolved—such as deconcentration, decentralization, or devolution (cf. Cheema and Rondinelli 1983)—because degree of authority devolved is part of what I seek to explain.

<sup>5</sup> Unlike other definitions of fiscal decentralization that collapse decentralization of revenues and expenditures, in this definition fiscal decentralization refers to revenues, whereas expenditures are part of administrative decentralization. This analytic separation makes it easier to evaluate the consequences of decentralization processes where the transfer of revenues and expenditures do not go hand in hand, allowing the disentangling of seemingly contradictory outcomes such as “centralization via decentralization” (see Wibbels 2004, 220–21).

<sup>6</sup> Although political decentralization and democratization can be mutually reinforcing, the two processes need to be distinguished analytically. For example, the return to free and fair elections at all levels of government after an authoritarian regime does not necessarily constitute a political decentralization policy. The transition to democracy may simply be reinstating the electoral norms and rules of the pre-authoritarian period, with no negotiation of a policy reform that specifically targets the subnational level. Similarly, if an electoral reform that is designed to augment political competition in the political system as a whole were to have the unintended consequence of increasing the power of subnational political actors, it cannot be considered a political decentralization measure because it was not planned, designed, or negotiated with the explicit goal of empowering subnational polities. To qualify as political decentralization, the reform in question must explicitly address the devolution of political authority or capacities to *subnational* polities.

By unpacking decentralization in this way, we see that, depending on their institutional design, decentralization policies can actually decrease the power of subnational officials with regard to the national executive. As shown in the following, the institutional design of decentralization policies is highly dependent on *when* those policies take place within the sequence of reforms. Political and fiscal decentralization policies that take place early in the sequence tend to increase the power of governors and mayors, whereas early administrative decentralization reforms tend to negatively affect their power.

### Territorial Interests of Bargaining Actors

The territorial interests of presidents, governors, and mayors are defined by the level of government (national, state, or municipal) and the characteristics of the territorial unit (e.g., rich or poor province, big city, or small town) they represent. Drawing from the literature on decentralization and in-depth interviews with national and subnational politicians and public officials,<sup>7</sup> I describe the set of preferences of national and subnational actors with regard to decentralization types.<sup>8</sup>

The national executive prefers administrative decentralization (A) to fiscal decentralization (F), which in turn is preferred to political decentralization (P), or  $A > F > P$ . The rationale of this ordering is that the national government seeks to divest itself of expenditure responsibilities first and foremost. Administrative decentralization is greatly preferred over the other two types of decentralization. As Garman, Haggard, and Willis (2001) say: "[W]e would expect the president to be more inclined to transfer responsibilities than the resources to meet them" (209). If the center is forced to choose between surrendering fiscal and political authority, it will choose to give away fiscal authority and to retain political control, which may serve to influence the expenditure decisions made by subnational officials.

The same reasoning applies to explain the reverse order of preferences of the subnational governments:  $P > F > A$ . Their preference, first and foremost, is political decentralization. If the president does not control the appointment and removal of governors and mayors, they can push forward the issues and concerns of their territorial units without fear of retaliation from above. If governors and mayors have to choose between fiscal and administrative decentralization, they will choose the transfer of revenues over responsibilities, particularly if the unions representing the public sectors to be

decentralized are large and strong. That is, subnational executives prefer political autonomy, money, and responsibilities, in that order.

### Sequences of Decentralization: Origins, Timing, and Mechanisms

The origins of the process of decentralization are important both theoretically and methodologically. On the one hand, the main argument of this article incorporates elements of path dependence, for which the issue of origins is crucial. On the other hand, the method of process tracing requires specifying when the process starts. Scholars have adopted different approaches to answer the question of when a path-dependent process starts, such as critical junctures (Collier and Collier 1991) or contingent events (Mahoney 2000). I define the origin of the decentralization process by the state *context* in which it takes place. As stated earlier, the contents of decentralization policies and their interaction with the broader political and economic systems are largely determined by the type of state they seek to reform. In Latin America, for example, in the context of the oligarchic states, decentralization reforms sought to consolidate or balance power among regional elites (Ansaldi 1992, 17). In the context of the developmental states, meanwhile, decentralization policies sought to strengthen certain regions to make them more adequate for private investment (González 1990, 75); whereas in the context of market-oriented states, decentralization policies largely sought to reduce the size of central governments. Of course, these were not the exclusive goals of decentralization reforms in each of these periods. Nonetheless, it is evident that in different historical periods the policies that transferred responsibilities, resources, or authority to subnational governments were part of state reform projects that had largely different overarching political and economic objectives. For this reason, when comparing across countries, the researcher should qualify the processes or sequences of decentralization by the type of state in which they take place, in order to assure the analytic equivalence of the compared policies.

For the purposes of this article, I focus on the process of decentralization that began with the transition from a "developmental" to a "public-goods" type of state (Block 1994). In Latin America, this was the transition from a *desarrollista* to a market-oriented type of state. During this period, decentralization policies were part of what became known as "second-generation" reforms (Camdessus 1999). Prior examples of centralization and decentralization existed in the region (see Eaton 2001, 2004; Montero and Samuels 2004, 14) and constitute the background against which the policies analyzed in this article took place. However, because they occurred in different types of state contexts, they form part of prior sequences of intergovernmental reforms.

Regarding the origin of the sequence analyzed here, although it is difficult to pin down the exact date when the *desarrollista* state ended in each case

<sup>7</sup> These are 86 in-depth interviews carried out in Argentina, Mexico, and Colombia during the summer of 1998, the spring of 1999, and between August 2000 and July 2001.

<sup>8</sup> The order of preferences of national and subnational officials helps to understand their position in the bargaining over different types of decentralization. However, I do not assume that these preferences are fixed throughout the entire decentralization process. Once the first decentralization policy has been implemented, its consequences on intergovernmental relations may reshape the bargaining actors' interests for subsequent rounds of reforms.

(Schneider 1999, 293), it is nonetheless possible to identify the first administration in each country that applied orthodox measures of economic adjustment and moved the state away from intervention in the economy. The market-oriented sequence of decentralization reforms starts with the first decentralization policy successfully implemented by the first administration that made the transition from a developmental economy and state toward a market-oriented economy and state.<sup>9</sup> Failed decentralization attempts do not constitute original moments because they do not have an impact on intergovernmental relations. They are analyzed as part of the process, as they may reflect on the distribution of power among bargaining actors, but do not constitute key transformative or original moments.

Using Skowronek's (1993, 9) terminology, we may conceive of intergovernmental relations as a layered structure of institutional action. Decentralization policies affect the fiscal, administrative, and political layers of intergovernmental relations. Rarely does a decentralization policy simultaneously affect all three intergovernmental layers (although it is possible). More often, different types of decentralization (as well as different policies within each type of decentralization) are negotiated and enacted at different points in time. Hence, the timing of each reform determines the particular sequence of decentralization that a given country undergoes. If the three types of decentralization defined previously take place (which is not theoretically necessary, but is a common occurrence), we can identify six sequences of decentralization according to the timing of the *first* decentralization policy within each intergovernmental layer. This does not mean that posterior decentralization policies in each layer do not happen or should be overlooked (see the analysis of empirical cases in the following section). However, the sequencing of the first decentralization policy in each layer is particularly important because it sets constraints on what is feasible in the remainder of the sequence and allows us to establish a basic model of the impact of different sequences of decentralization reforms on the intergovernmental balance of power.

The level of government whose territorial interests prevail at the origin of the decentralization process is likely to dictate the first type of decentralization. The first round of decentralization, in turn, produces policy feedback effects that account for the order and characteristics of the reforms that follow. If subnational interests prevail in the first round of negotiations, political decentralization is likely to happen first. Political

decentralization is likely to produce a policy ratchet effect (Huber and Stephens 2001, 10): a group of supporters who will continue to push in the direction of further decentralization. The formation of associations of governors, mayors, or similar instances of coordination of subnational politicians is an example of such policy ratchet effect. Lobbying through these associations, governors and mayors will enhance their power and capacities for the next rounds of decentralization. Even if this coordination mechanism is not in place, governors and mayors will find themselves in a better position to advance their preferences in the second round of reforms because they will enjoy greater political autonomy from the national executive. The president, moreover, may become more dependent on elected governors and mayors for the mobilization of votes in national elections. Thus, in the second round of decentralization, governors and mayors will most likely demand fiscal decentralization and influence its terms. Administrative decentralization, which after fiscal decentralization is likely to follow to compensate for the previous decentralization of resources (Haggard 1998, 217), will be the last type of reform. Administrative decentralization will therefore be funded and will not have a negative impact on the power of governors and mayors. The final outcome of this trajectory of decentralization ( $P \rightarrow F \rightarrow A$ ) that conforms to the preferences of the subnational officials is likely to be a high degree of autonomy for governors and mayors with respect to the president (see Table 1). I show below that Colombia followed this path from 1986 to 1994.

If, instead, national interests prevail at the beginning of the process, administrative decentralization is likely to occur first. If fiscal resources do not accompany the transfer of responsibilities, the national executive will strengthen its power vis-à-vis subnational officials, who will become more dependent on transfers from the center. If the process of decentralization continues, the president will choose fiscal over political decentralization. But due to a power reproduction mechanism (Stinchcombe 1968, 117–18), the national executive will control the timing, pace, and contents of the reform. Governors and mayors, under the fiscal strain of the first round of unfunded administrative decentralization, will be in no position to reject those terms set by the center—unless exogenous circumstances were to change their relative power vis-à-vis the president. Following this trajectory, political decentralization, if it happens, will be the third type of reform. The outcome of this trajectory of reforms ( $A \rightarrow F \rightarrow P$ ) that conforms to the preferences of the national executive is likely to be little or no change in the redistribution of power to the subnational authorities. I show in the following that Argentina followed this path of reforms from 1978 to 1994.

It is also possible that exogenous changes (such as midterm elections, a context of fiscal expansion, fiscal crisis, or a process of democratization) could produce reversals on the distribution of power between national and subnational executives once the process of decentralization has started. This would lead to the alternative sequences  $P \rightarrow A \rightarrow F$  and  $A \rightarrow P \rightarrow F$ . In the first

<sup>9</sup> In the countries analyzed in this article these administrations were the military governments of Jorge R. Videla in Argentina (1976–1981) and João Figueiredo in Brazil (1979–1985) and the presidencies of Belisario Betancur in Colombia (1982–1986) and Miguel de la Madrid in Mexico (1982–1988). In most of Latin America, the transition from state interventionism to free-market economies was the response to the economic troubles unleashed by the debt crisis of the early 1980s (albeit not in Argentina and Chile, where the move to free-market economies preceded the foreign debt crisis). Subsequent administrations applied both orthodox and heterodox economic policies, but the move away from developmentalism had already taken place (see Weyland 2002, 72, 77–81).

**TABLE 1. Sequences of Decentralization and Their Effects on the Intergovernmental Balance of Power**

Prevailing Interests in First Move	First Type of Decentralization	Type of Feedback Mechanisms	Second Type of Decentralization	Third Type of Decentralization	Degree of Change in the Intergovernmental Balance of Power <sup>a</sup>
Subnational	→ Political decentralization	Self-reinforcing	→ Fiscal decentralization	→ Administrative decentralization	High
National	→ Administrative decentralization	Self-reinforcing	→ Fiscal decentralization	→ Political decentralization	Low
Subnational	→ Political decentralization	Reactive	→ Administrative decentralization	→ Fiscal decentralization	Medium/low
National	→ Administrative decentralization	Reactive	→ Political decentralization	→ Fiscal decentralization	Medium
Tie	→ Fiscal decentralization	Reactive	→ Administrative decentralization	→ Political decentralization	Medium/low
Tie	→ Fiscal decentralization	Self-reinforcing	→ Political decentralization	→ Administrative decentralization	High

<sup>a</sup> A positive direction of change in the intergovernmental balance of power reflects a move toward greater autonomy of subnational officials vis-à-vis national officials. Thus, a "high" value in the degree of change of the intergovernmental balance of power corresponds to a higher degree of autonomy for governors and mayors, whereas a "low" value indicates that the degree of autonomy of subnational officials has remained practically unchanged.

scenario, subnational interests prevail at the beginning of the sequence, triggering political decentralization. However, reactive mechanisms (such as a fiscal crisis that undermines subnational demands for fiscal decentralization) lead to a prevalence of national interests in the second round and, thus, to administrative decentralization. The last stage (should it happen) is fiscal decentralization. This trajectory may be disastrous for subnational officials if administrative decentralization is unfunded. If they are granted political autonomy, and soon after that they receive unfunded responsibilities, their subnational constituencies will blame them for poor performance. Most likely, this trajectory will lead to a low change in balance of power. If administrative decentralization is instead funded, this trajectory may lead to a medium degree of change in intergovernmental balance of power. In the second scenario, national interests prevail at the beginning of the sequence, but reactive mechanisms (such as a process of democratization that undermines centralized power) afford subnational executives the possibility of pushing political decentralization forward in the second round. In this situation, subnational actors (due to the political power they now have) are in a better position to set the terms of fiscal decentralization. The overall outcome of this trajectory would be a shift in the balance of power in favor of subnational authorities, but not as significant as in the first aforementioned trajectory (P → F → A).

Finally, we could also conceive of a tie between national and subnational interests at the outset of the reform process, such that no side is capable of achieving its most preferred outcome. In this situation, either the status quo will prevail or bargaining actors will compromise in their second most preferred outcome: fiscal decentralization. If this happens, the way in which the sequence continues will depend on the effects of this reform on the relative power of national and subnational executives. If the national executive prevails, administrative decentralization should follow, with political decentralization happening last (F → A → P). This trajectory should lead to a medium to low change in balance of power. The crucial issue here is the time lag between the first and second rounds of reforms. If subnational officials receive money without strings, and they can apply it to strengthen their support base and popularity for a considerable amount of time before they receive new responsibilities, this trajectory may lead to a medium increase in balance of power, even if political decentralization only takes place at the end of the trajectory. In contrast, if money and responsibilities are decentralized practically at the same time, this means that subnational officials are receiving new responsibilities without political autonomy. The impact of decentralization of funded responsibilities on balance of power will then be highly dependent on how successful the subnational governments are in efficiently delivering the newly transferred services. Considering, however, that the subnational officials are probably more accountable to the national executive than to their local constituencies (recall that political decentralization does not take place until the end of



the trajectory), administrative performance will likely be poor, and the result will be small change in intergovernmental balance of power. Alternatively, if after a tie of territorial interests and a first round of fiscal decentralization subnational executives prevail, political decentralization should be next, with administrative decentralization taking place last ( $F \rightarrow P \rightarrow A$ ). This sequence should lead to a high change in balance of power in favor of the subnational governments. In this type of trajectory, subnational governments gain fiscal capacities, then political autonomy; and last, they receive administrative responsibilities. The first two moves in this sequence should allow subnational authorities to build strongholds of supporters (because they have the resources to do so) and to win elections. Once this happens, they gain greater autonomy from the national executive, as illustrated in the bottom line of Table 1.

The assumptions thus far have been that the three types of decentralization take place and that a sequence among them can be established. Moreover, I have only taken into account the first successfully implemented policy within each type of decentralization and the first cycle of decentralization, which ends once the three types of reforms have occurred. Decentralization processes, however, could evolve differently in reality. Only one or two types of reforms could occur, the timing of policies could overlap, and successive reforms within each layer could affect those that follow. Some of these complexities will be revealed in the analysis of the cases in the following sections. Nonetheless, as long as at least two types of devolution of authority and two implementation moments can be identified, the proposed sequential reasoning could be modified accordingly and applied to cases and sequences that follow different patterns.

## EVOLUTION OF INTERGOVERNMENTAL BALANCE OF POWER IN LATIN AMERICA

Intergovernmental balance of power is defined as the *relative* power or degree of autonomy of subnational officials with regard to national officials. Intergovernmental power is dependent on (1) economic resources, which enhance the capacity of political actors to pursue their desired courses of action; (2) legal authority, which sets the institutional limit that economic resources can reach; and (3) organizational capacities, which facilitate coordination at each level of government.

Because this article is concerned with the effects of decentralization on the evolution of balance of power, in operationalizing this concept, the focus is precisely on those dimensions of intergovernmental power susceptible to change due to the implementation of decentralization policies. Building on the works of Stepan (2004) and Samuels and Mainwaring (2004), intergovernmental balance of power is operationalized in five dimensions: (1) the subnational share of revenues, which measures the percentage of public money collected by subnational governments (provin-

cial and municipal); (2) the subnational share of expenditures, which measures the percentage of public money allocated by subnational governments; (3) the policymaking authority, which measures the degree of autonomy of subnational officials to design, evaluate, and decide on issues concerning a specific policy area; (4) the type of appointment of subnational officials, which records whether governors and mayors are elected or appointed; and (5) the territorial representation of interests in the national legislatures, which reports the average degree of overrepresentation of the subnational units in the lower and upper chambers of congress. If decentralization reforms were always to increase the power of subnational officials, we would observe a positive change in all these indicators. If, however, it is possible for decentralization not to increase the power of subnational officials, we would expect some of these indicators to decrease in value or to remain unchanged.

The remainder of this section compares the absolute levels of decentralization before and after decentralization and analyzes the degree of change in intergovernmental balance of power in the four largest countries of Latin America—the region that took the lead in the implementation of decentralization reforms (Camdessus 1999). Several commonalities make Argentina, Brazil, Colombia, and Mexico suitable countries for comparison. First, due to their size, it is safe to assume that relationships between center and periphery are contentious and that issues of decentralization are politically relevant. Second, they all underwent similar decentralization policies, although with different impact on the intergovernmental distribution of power. Third, they all have similar structures of government, with three tiers of government and bicameral national legislatures. Finally, differences among the cases allow for controls to the main argument. Although Argentina, Brazil, and Mexico are federal countries, Colombia is a unitary country; and although Argentina, Brazil, and Colombia have decentralized party systems, Mexico has a centralized one.

In Table 2, the first two columns within each country measure the *absolute level of decentralization* (term used as to describe the state of being of the fiscal, administrative, and political systems), and the third column measures the *relative degree of change in the intergovernmental balance of power*. Along the fiscal dimensions, the *subnational share of revenues* (SSR) decreased in Argentina and increased in the other three countries; whereas *subnational share of expenditures* (SSE) increased in the four countries. At the beginning of the period, Argentina and Brazil had the highest absolute levels of fiscal decentralization, in terms of both revenues and expenditures, followed by Colombia and Mexico, in that order. By the end of the period a different pattern emerged. Brazil continued to be fiscally the most decentralized, but now Colombia was second, and in SSR Mexico surpassed Argentina, which had the lowest collection of subnational revenues and the highest fiscal vertical imbalance of the four countries. Relative to the initial conditions, Mexico was the country whose fiscal structure changed the most, followed

TABLE 2. Absolute Level of Decentralization and Evolution of Balance of Power in Argentina, Brazil, Colombia, and Mexico, 1978–1999

Variable	Argentina			Brazil			Colombia			Mexico		
	Prior Dec.	After Dec.	Change	Prior Dec.	After Dec.	Change	Prior Dec.	After Dec.	Change	Prior Dec.	After Dec.	Change
SSR	21% (1983) 3	19% (1999) 1		25% (1980) 4	33% (1995) 4		18% (1980) 2	28% (1995) 3		9% (1982) 1	21% (1997) 2	
SSR level												
SSR change <sup>a</sup>			-8% 1									
SSR ch. rank.						29% 2			56% 3			122% 4
SSE	34% (1978) 4	41% (1999) 3		32% (1980) 3	44% (1995) 4		28% (1978) 2	40% (1995) 2		18% (1978) 1	29% (1999) 1	
SSE level												
SSE change <sup>b</sup>			21% 1			38% 2			43% 3			61% 4
SSE ch. rank.												
PMA <sup>c</sup>												
Curricula	C	C	0	C	S	0.5	N	N	0	N	N	0
Teachers' training	C	C	0	C	S	0.5	N	C	0.5	N	C	0.5
Evaluation	C	C	0	C	S	0.5	N	C	0.5	N	C	0
School management	C	S	0.5	C	S	0.5	S	S	0	C	S	0.5
Hire, fire, relocation	C	S	0.5	C	S	0.5	N	S	1	N	S	1
Salary	C	S	0.5	C	S	0.5	N	C	0.5	N	C	0.5
PMA level	(1978) 3.5	(1994) 3		(1982) 3.5	(1995) 4		(1982) 2	(1994) 2		(1978) 1	(1994) 1	
PMA change			1.5 1			3			2.5			2.5
PMA ch rank.						4			2.5			2.5
ASO <sup>d</sup>												
Governors	E	E	0	E	E	0	A	E	1	A/E	E	0.5
Mayors	A/E (1983) 3	E (1996) 2.5	0.5	E (1982) 4	E (1994) 2.5	0	A (1982) 1	E (1991) 2.5	1	A/E (1982) 2	E (1994) 2.5	0.5
ASO level												
ASO change			0.5 2			0			2			1
ASO ch. rank.						1			4			3



by Colombia and Brazil. Argentina was the country that changed the least, even experiencing a negative change in SSR.

Regarding the administration of social services, the dimension *policymaking authority* (PMA) is applied to the educational sector. The selection of education over other policy sectors responds to several reasons. First, in most countries, education was the first public sector to be decentralized, influencing the pace and characteristics of decentralization in other areas. Second, education is the largest public sector in these countries, in terms of both fiscal and human resources. The transfer of education carries, therefore, significant fiscal and administrative consequences for states and municipalities. Finally, the education sector has often strong and large unions. This makes decentralization of education politically crucial for national and subnational executives, who have to negotiate with the teachers' unions. The six indicators taken into account within this dimension were authority over the curricula; responsibility for training teachers; responsibility for evaluation of the educational system; management of schools; authority over hiring, firing, and relocation of teachers; and authority over salaries. At the beginning of the period, the countries can be paired in terms of the distribution of responsibilities among levels of government: Argentina and Brazil were the most decentralized, and Mexico and Colombia were the most centralized. By the end of the period, the ordering of the countries is similar, but Brazil experienced a greater degree of devolution of authority to subnational authorities than Argentina. Whereas in 1982 the Brazilian states and the federal government shared responsibilities along all of the educational indicators considered (Tavares de Almeida 1995, 20, 27), by the mid-1990s all of these issues lay in the hands of governors, mayors, or school directors (Burki, Perry, and Dillinger 1999, 71). Mexico and Colombia follow Brazil in the degree of change in PMA. In Mexico, all issues of public education management were in the hands of the federal government in 1978 (with the sole exception of the management of school buildings). In 1992, after the signing of a decentralization agreement, authority over the curricula and evaluation of the system remained at the federal level, but all other issues were decided on by the subnational level or jointly by both levels of government. The situation in the education sector in Colombia by the early 1980s was similar to that in Mexico: all responsibilities resting with the national government, with the exception of the maintenance of schools. But after the decentralization of education in 1992 and 1993, all educational issues became matters of state authority (with the sole exception of the design of the curricula, which remained in the hands of the central government). In Argentina, the situation was different. By the mid-1970s the Argentine provinces managed half of the public primary and secondary schools, which meant that all responsibilities concerning the public educational system had historically been shared by the federal and provincial governments. Decentralization of primary and secondary schools (in 1978 and 1992, respectively) did not change

the distribution of formal authority among levels of government. This change only came about when a new federal education law was passed in 1993 and some educational issues became the sole domain of the provinces (Corrales 2004). As can be seen in Table 2, in terms of PMA, Brazil experienced the most change, followed by Mexico, Colombia, and Argentina, in that order.

Two dimensions account for the distribution of power in the political arena. The first one is the *appointment of subnational officials* (ASO). Along this dimension, and because of its starting point, Colombia is the country that changed the most. Prior to decentralization, mayors and governors were appointed; their offices became popularly elected in 1988 and 1991, respectively. Mexico follows Colombia in degree of ASO change. In Mexico there were elections for subnational offices (with the exception of Mexico City's mayor), but they were not competitive. It was not until the mid-1990s that elections for mayors and governors became (by and large) competitive in Mexico. Next is Argentina. The office of the mayor of the city of Buenos Aires was politically decentralized in 1994, but the other mayors and governors had historically been popularly elected. Finally, ASO remained constant in Brazil throughout the period of reforms.

The second political dimension is the *territorial representation of interests* (TRI). In this dimension, overrepresentation coefficients report the degree of deviation from the principle "one citizen, one vote." A coefficient value of 1 indicates proportionality between seats and population. If the overrepresentation coefficient is higher than 1, it means that in some subnational units the "cost" of electing a deputy or a senator is lower than in others. In Stepan's (2000) words, the higher the coefficient the more "demos-constraining" these Senates or Houses are. The higher the overrepresentation coefficients, the easier it is for some of the deputies and senators to represent the territorial interests of their subnational units and constituencies, instead of the interests of the political majority. Brazil and Colombia are the countries that experienced the highest degrees of change in overrepresentation in either one or both of their chambers. In Brazil, the creation of two new states (Mato Grosso do Sul and Tocantins) and changes introduced in the 1988 constitutional reform meant that between 1962 and 1995, the degree of overrepresentation in the lower chamber increased from an average of 1.51 to 1.92. The changes were even more drastic in the Senate, where the allocation of seats to previously unrepresented and relatively small subnational units meant that the average degree of overrepresentation increased from 2.66 in 1978 to 3.94 in 1995. In Colombia, as a consequence of the changes introduced in the 1991 constitutional reform and the allocation of seats to 7 previously unrepresented departments, the average degree of overrepresentation of subnational units in the lower chamber increased from 1.17 in 1982 to 2.73 in 1994. The Senate, whose seats were distributed among 23 departments according to population prior to 1991, was



transformed after the constitutional reform into a proportionally representative chamber of 100 members chosen from a single national district. In Argentina and Mexico, the degrees of overrepresentation in the lower and upper chambers practically did not change. Argentina had a high degree of overrepresentation of subnational units in the Senate throughout the period of decentralization reforms (3.15 in 1983 and 3.40 in 1995 after the incorporation of Tierra del Fuego) and had a moderately high degree of overrepresentation in the lower chamber (1.94 in 1983 and 1.85 in 1995). Mexico had a similar degree of overrepresentation in the Senate as had Argentina in its lower chamber (1.96)—and this stayed the same throughout the period. In the Mexican lower chamber representation was proportional (1.00). Hence, in terms of degree of change in TRI, Brazil experienced the most, followed in decreasing order by Colombia, Argentina, and Mexico.

In summary, an overview of the position of each country along each one of the five variables reveals that prior to decentralization reforms Argentina and Brazil had the highest absolute levels of decentralization, whereas Mexico and Colombia had the lowest. This corresponds to what we know about how federalism and intergovernmental relations have historically evolved in these countries (Gibson and Calvo 2000, Gibson and Falleti 2004, Samuels 2003). Nevertheless, if we look at the overall change in balance of power that occurred after decentralization policies were implemented, we find that although Colombia, Brazil, and Mexico experienced significant shifts in balance of power in favor of the subnational authorities, the intergovernmental balance of power in Argentina stayed practically the same throughout the period. At one extreme, Colombia saw its subnational share of revenues and expenditures increase by a ratio of 0.56 and 0.43, respectively, its governors and mayors gain significant authority in the administration of public education, its president lose the authority to appoint subnational officials, and the territorial overrepresentation in its chamber of deputies over double. At the other extreme, Argentina saw virtually no change in intergovernmental balance of power. The share of revenues decreased whereas the share of expenditures increased, augmenting the fiscal vertical imbalance in subnational accounts. Administrative decentralization did not confer new capacities to subnational executives until 1993. Political decentralization, although beneficial to the city of Buenos Aires, did not have an impact on the rest of the provinces. As described in a World Bank report, "Argentina is arguably one of the most decentralized countries in [Latin America] but has essentially the same political and fiscal structure it had before the military intervened in 1976. In contrast, Colombia has radically increased the power and responsibilities of subnational units of government" (Burki, Perry, and Dillinger 1999, 11). Why, despite the implementation of decentralization reforms, did Argentina's fiscal and political intergovernmental structure remain unchanged, while Colombia's fiscal and political intergovernmental relations changed so radically?

## ALTERNATIVE EXPLANATIONS

One possible explanation is that Argentina did not devolve more power to governors and mayors because of its high initial level of absolute decentralization. In other words, it could be argued that there is an upper limit on the degree of change that decentralization can bring about in intergovernmental relations, or a threshold of devolution of power below which a country cannot fall. However, the evolution of intergovernmental balance of power in Brazil challenges this interpretation. Brazil started the period with a fiscal, administrative, and political structure as decentralized as that one of Argentina. However, by the end of the period, decentralization policies in Brazil (in the fiscal, administrative, and political spheres) had produced significant changes to the intergovernmental structure such that more power was devolved to governors and mayors. This was evident along the subnational share of expenditures and revenues, the distribution of policy-making authority, and the political reforms introduced in the 1988 constitution. Interestingly, Argentina underwent similar policies in the administrative, fiscal, and political arenas, but their impact in augmenting the power of governors and mayors was far more limited.

The second explanation draws from Riker's (1964) theory of federalism and argues that the degree of autonomy of subnational officials after the implementation of decentralization reforms can be explained by reference to the internal structure of the political parties (Garman, Haggard, and Willis 2001). This argument states that if—given certain electoral and nomination procedures—national legislators are more accountable to the national executive, they will tend to push for more centralization of authority in the design of and bargaining over decentralization reforms. If instead the national legislators are accountable to subnational officials, they will press for further decentralization of power in designing these policies. This explanation successfully accounts for the absolute levels of decentralization before and after the reforms. However, it cannot account for the degree of change in intergovernmental relations. Argentina has a decentralized political party system, with national legislators accountable (mostly) to subnational authorities (Eaton 2002; Jones et al. 2002). Nonetheless, Argentina is the country where intergovernmental balance of power evolved the least. Mexico, on the other hand, has a centralized party system, but its intergovernmental balance of power changed considerably once decentralization measures were undertaken.

Finally, it could also be argued that the degree of change in intergovernmental relations that decentralization brings about is dependent on the constitutional type of government. Because federal constitutions confer autonomy to subnational units, this guarantee should lead to higher levels of devolution of power than experienced in unitary countries (Dahl 1986). My cases show the opposite to be true. In Colombia, a unitary country, decentralization had the most significant impact on the evolution of intergovernmental balance of power. In Argentina, a federal republic,

decentralization had the least significant impact on intergovernmental balance of power.<sup>10</sup>

## THE SEQUENTIAL THEORY OF DECENTRALIZATION APPLIED

To illustrate the range of the proposed theory, this section traces the trajectories of decentralization in the two extreme cases: Colombia and Argentina. From the late 1970s to the mid-1990s, Colombia and Argentina both underwent processes of decentralization that accompanied the movement from state-led to free-market economies. In both cases, fiscal, administrative, and political decentralization reforms took place, and decentralization was pursued under the pretense of strengthening the subnational units. In spite of these similarities, the processes of decentralization and the consequences they brought about for intergovernmental relations were radically different, as described previously. These differences can be appreciated more fully by analyzing the evolution of the first cycle of political, fiscal, and administrative reforms. In what follows, I argue that the different outcomes for intergovernmental balance of power are less a result of the particulars of individual policy reforms than a product of the *evolution* of such reforms and of the type of actors they empower along the way.

### Colombia: The Subnational Path to Decentralization

In 1986, by initiative of Conservative President Belisario Betancur (1982–1986), the younger and less entrenched factions of the two traditional parties in congress (Liberal and Conservative) passed a constitutional amendment for the popular election of mayors. This law changed one hundred years of intergovernmental relations. Since 1886, the president had appointed the governors, who in turn appointed the mayors. President Betancur explained in the following terms his support for this measure:

I had the conviction; I had the obsession that the community should be closer to their representatives. I knew that as long as the community was closer to the rulers, those rulers would feel more stimulated, with greater support to govern . . . If popularly elected, mayors would be freer and more efficient. (Betancur, Belisario, interview by author, Bogotá, March 28, 2001)

However, the decision to popularly elect the mayors did not result solely from the president's political convictions. According to O'Neill (1999, 2003), presidents are more likely to implement political decentralization when the prospects of their parties winning future national elections are bleak, while at the same time strong pockets of support exist throughout the country that would win them elected positions at the

subnational level. Although O'Neill's (2003) theory is compelling, she also notes "it would be absurd to ignore the importance of context-specific factors that affected decisions to decentralize" (1070). My contention is that when other types of decentralization are considered, those context-specific factors help to account not only for the timing of decentralization but also, and more importantly, for the type and content of the policy first implemented. In the case of Colombia, the social mobilizations against the shortcomings of the developmental state help to explain why and *how* decentralization came about. They reveal the presence of territorial subnational interests in the coalition that pushed decentralization forward, a presence that has been largely overlooked.

During the 1960s and 1970s, the planning and implementation of developmental policies had been transferred to parastatal institutions, relatively autonomous agencies attached to central offices and ministries. They were equipped with significant financial resources and were designed to operate in a cost-recovery basis and on a nationwide scale. These agencies supplanted the role of local government in areas such as urban planning, housing, health, education, and the provision of services such as electricity, water, and sewage. The coverage was not uniform, however. Large municipalities kept the management of more responsibilities, and peripheral, poorer regions were left largely unattended. The parastatal agencies tended to focus more heavily on the regions prone to private investment, which created profound regional inequalities (Collins 1988, 426–27; Maldonado 2000, 72). Moreover, local government expenditures had dropped from 18% of total expenditures in 1967 to 14% in 1978 and were concentrated in the largest cities. In 1979, the three largest municipalities (Bogotá, Medellín, and Cali) absorbed 72% of the total local government expenditures, even though they accounted for 26% of the population. After the rest of the departmental capitals were considered, only 13% was left to be spent in more than nine hundred remaining municipalities, where over 35% of the population lived (Collins 1988, 426; DNP and PNUD 1998, 39; Nickson 1995, 146). This created ample discontent among the inhabitants of the poorer regions.

Between 1971 and 1985, over two hundred civic strikes (*paros cívicos*) took place. These strikes "involved the total or partial paralysis of social and economic activity in urban centers and/or regions as a means of pressing the state to accede to demands" (Collins 1988, 425). Sixty percent of the strikes were related to problems in the delivery of electricity, water, and sewage; 9%, to problems with roads; 6%, to problems in education; and 5%, to ecological problems (Velásquez 1995, 246). The majority of these strikes occurred in midsized municipalities (with 10 to 50 thousand people) in the country's peripheral regions, particularly in the departments of the Atlantic coast (Maldonado 2000, 73). Broad sectors of the population participated in these strikes, voicing the territorial interests of the underdeveloped regions. As Jaime Castro, former mayor and member of the 1991 constitutional convention, said:

<sup>10</sup> As Escobar-Lemmon (2001, 27) writes "while state structure may explain the initial level of decentralization in a country, with federal cases being more decentralized, it does little to explain changes within a country over time."

The civic strikes had become the mechanisms of protest of *la provincia* [the interior] in relation to the central government. The civic strikes brought to the forefront the fact that it was necessary to strengthen the municipalities and departments. . . . They continued to happen after the popular election of mayors, but I would say that thanks to decentralization civic strikes have now disappeared. (Castro, Jaime, interview by author, Bogotá, March 29, 2001)

The civic strikes brought local government to the center of the political scene in several ways. First, they pointed out the deficiencies of the parastatal agencies and the local administrations in delivering public services. The national executive paid close attention to this problem. In 1980, a team of economic experts was formed to study how to improve the system of intergovernmental finances. Richard Bird led this team, whose findings and recommendations were published a year later (*Misión de Finanzas Intergubernamentales* 1981). When the next president, Belisario Betancur, was confronted with increasing economic problems and steadily declining municipal and departmental revenues, he passed an emergency plan that included some of these recommendations. Law 14 of 1983 sought to strengthen the collection of taxes in departments and municipalities. Departments were given a new tax on automobiles and the authority to update and simplify their existing taxes, whereas municipalities could modernize their tax bases—important for property taxes—and determine within certain parameters their own level of industry and commerce tax (Ocampo Gaviria and Perry Rubio 1983). This fiscal measure halted the trend of declining municipal and departmental revenues and, although its overall impact on the distribution of resources among levels of government was negligible (Wiesner Durán 1992, 117–29), it revealed the importance of subnational pressures. Second, the civic strikes were signs that the old system of handpicked mayors was coming to an end. Local bosses and traditional clientelist practices had proved inadequate in alleviating popular discontent. The political appointment of mayors had led to a system in which mayors were dependent on the legislator, the governor, or the president—whoever was politically responsible for their appointment—and only accountable to them. There were frequent changes of local administrations and corruption was pervasive (Gaitán Pavía and Moreno Ospina 1992, 150–51). Very often mayors were not native to the town they governed. A number of these became known as “professional mayors,” who “would travel around all the municipalities of one department until they were discredited in all of them” (Osorio, Luis Camilo, interview by author, Bogotá, July 30, 1998). Finally, the strikes showed that there were locally based citizens who were demanding accountability and better services in their municipalities. These were broad nonpartisan civic coalitions that helped to put municipal democratization on the agenda. Decentralization in Colombia was thereby initiated from below. It was fueled by the protests of the local communities. When the president proposed and the national legislators passed the political decentral-

ization reform of 1986, they were responding to those subnational demands and interests voiced in the civic strikes.

What were the consequences of the direct election of mayors? The immediate result was a decline in the number of civic strikes. There were 51 strikes in 1987, 35 in 1988, and only 19 in 1989 (Correa Henao 1994, 48–54). Former guerrilla members were incorporated into the legal political system. In some cities and regions, the grip of traditional caciques and local bosses loosened, and competition for public office presented them with new challenges they had never had to face in the past (Angell, Lowden, and Thorp 2001; Velásquez 1995). The direct election of mayors also produced two major policy ratchet effects: (1) *incrementalism* in the political sphere and (2) *coordination* among subnational authorities.

The 1986 decentralization reform created an impulse to further develop political decentralization, and this impulse would prove difficult to reverse. At the beginning of 1991, a constitutional assembly convened in Bogotá. The assembly, in session from February to July 1991, was organized into five committees. The second committee was responsible for territorial organization. Two of the main issues discussed in this committee were the popular election of governors and the degree of autonomy to be conferred to the intermediate level of government. The assembly was split between the so-called *departamentalistas*, who were in favor of the popular election of governors, and the *municipalistas*, who opposed it. However, against a backdrop of popularly elected mayors, the election of governors came to be seen as an inevitable next reform, even by the *municipalistas*. As one of them said, “The popular election of governors appeared to be a complement to the popular election of mayors. It was the next step” (Castro, Jaime, interview by author, Bogotá, March 29, 2001).

Political decentralization in 1986 also fostered coordination among the beneficiaries of the reform. It created a group of followers interested in deepening decentralization. The clearest manifestation of such an effect was the creation of an association of mayors. In 1988, the first cohort of elected mayors created the Colombian Federation of Municipalities (*Federación Colombiana de Municipios*, or FCM). As expressed in its statutes, the mission of the association is: “[T]o represent the collective interests of the municipalities, to lead and support the development of the municipal management, and to promote the deepening of decentralization” (FCM 1991). In 1991, FCM was very active in lobbying conventionalists for the extension of the mayors’ tenure from 2 to 3 years, for the recognition of municipal autonomy in the constitution, and for the transfer of more fiscal resources to municipalities (*El Tiempo*, Bogotá, 23 February and 23 March 1991). Despite the reluctance of the national executive, all these reforms were approved and political and fiscal decentralization were deepened as a result.

Although previous measures in the direction of transfer of revenues and expenditures to subnational governments had been taken, their impact in the

distribution of resources between levels of government was negligible.<sup>11</sup> However, after the first round of political decentralization and the creation of FCM, a major fiscal decentralization reform was incorporated in the 1991 constitution. Article 357 of the new constitution established that the transfers to municipalities would increase from a level of 14% of the current national income in 1993 to 22% in 2002. This reform expanded not only the rate but also the base of the automatic transfers, which included thereafter both tax and non-tax revenues. As a consequence, the total transfers to subnational governments (both departments and municipalities) passed from 38% to 52% of the current national income between 1991 and 1997 (Vargas González and Sarmiento Gómez 1997, 33).

The administrative counterpart to fiscal decentralization came about in 1993. The initial impetus to pass this reform came from the national executive, which was eager to establish a new distribution of responsibilities among levels of government as a means to cut the double spending and the deficit that fiscal decentralization had introduced in 1991. The national executive sent the administrative decentralization bill proposal to congress in mid-1992. It took 1 year from the presentation of the bill proposal until the final approval of Law 60 in August 1993. Law 60 became to be known as the "framework law" of administrative decentralization. It ruled on the distribution of responsibilities among levels of government regarding education, health, housing, and water and sewage. It was the result of compromises made by the national executive, the representatives of states and municipalities, and the national teachers' union. The national minister of education mediated between the interests of the ministry of economy and the department of national planning, who wanted to take decentralization of education to the municipal level, and those of the union, which was opposed to decentralization, particularly toward the municipal level. With the agreement of subnational representatives, the compromise reached between the union and the national government was that decentralization of education would take place toward the intermediate level of government, *with funds* guaranteed from the national level (Angell, Lowden, and Thorp 2001, 178). The departments thereby became responsible for paying and training teachers. They could also give vouchers to students with special needs. The municipalities were responsible for investing in the construction and maintenance of school buildings. Together, departments and municipalities were responsible for managing the educational services of preschool, primary school, secondary school, and high school. The national level retained jurisdiction over curricula and general educational guidelines, and the three levels shared responsibility for the evaluation of the educational system. Apart from the distribution of responsibilities among

levels of government, the law also established the distribution of resources among the subnational units and the creation of committees (*comisiones veedoras*) at both the departmental and the municipal levels to ensure that the transfers were properly allocated according to the law. It also granted FCM 0.01% of the total transfers to the municipalities "for the promotion and representation of all its members... the districts and municipalities" (Article 37, Law 60). Administrative decentralization was thus favorable to subnational authorities, regarding both policymaking and fiscal capacities. This was largely due to the fact that political and fiscal decentralization had already taken place and subnational interests were effectively represented by the time administrative decentralization came about.

The process of decentralization in Colombia followed a sequence of reforms that conformed to the preferences of subnational actors. Political autonomy was devolved first, followed by resources, and finally by responsibilities. The decision to popularly elect mayors in Colombia had self-reinforcing effects on the next rounds of political, fiscal, and administrative reforms. It produced coordination among subnational authorities that led to fiscal decentralization and deepened political decentralization through the extension of the mayor's mandate and the recognition of municipal autonomy in the national constitution. It also produced a sense of incrementalism in the political elite that allowed for the approval of the popular election of governors. Administrative decentralization was the last, almost residual, type of reform. It was pushed through by the national executive. However, owing to the sequence of previous decentralization reforms, subnational actors and the teachers' union were able to get the guarantee that the fiscal resources necessary to afford the costs of the transferred services would also be transferred. As a result, this measure did not have a negative effect on the degree of autonomy of subnational executives with regard to the national government. As is evident in Table 2, this first cycle of political, fiscal, and administrative decentralization in Colombia empowered subnational executives.

### Argentina: The National Path to Decentralization

Unlike the case of Colombia, Argentina's path of decentralization conformed to the preferences of the national executive. After the move away from developmentalism, the process of decentralization started with an administrative reform in 1978. It was followed by fiscal decentralization in 1988, and finally by political decentralization in 1994.<sup>12</sup>

On June 5, 1978, the national military junta passed two decrees transferring all national preschools and primary schools to the provinces, the city of Buenos

<sup>11</sup> For this reason, they do not count as prior instances of decentralization. These reforms were Law 14 of 1983—described earlier—, Law 12 of 1986, Law 29 of 1989, and Law 10 of 1990 (for a complete list and description of measures, see Gaitán Pavia and Moreno Ospina 1992, 283–94).

<sup>12</sup> Other decentralizing and centralizing reforms followed (see Eaton and Dickovick 2004). I focus here on the first cycle of decentralization, which ends once the three types of decentralization (fiscal, administrative, and political) have taken place.

Aires, and the territory of Tierra del Fuego. Retroactive to January 1, approximately 6,500 schools, 65,000 public employees, and 900,000 students (about one third of the primary public education system) were transferred to the provincial administrations. No revenues or fiscal capacities were transferred with the schools, and yet the transfer had a cost of 207 billion pesos—equivalent to 20% of the total national transfers (FIEL 1993, 148).

National interests prevailed in this first round of decentralization. In the context of an authoritarian regime, the national executive was able to impose on the provinces its most preferred outcome: administrative decentralization. The central government was interested in administrative decentralization for several reasons. First, they saw the provinces as enclaves of conservatism, in which future right wing political parties could develop. Second, the central government was interested in cutting the size of the federal bureaucracy and the national deficit, in the spirit of a neoliberal program of government (Novick de Senén González 1995, 138). Third, an increase in provincial revenues—which rose from 0.88% in 1976 to 1.56% of the GDP in 1977 (Kisilevsky 1998, 55)—established a favorable environment to transfer expenditures without resources. A report by the national ministry of education gave the following account of conditions before the 1978 transfer:

At the end of 1977, the national minister of economy [José Martínez de Hoz] considered that there had been an increase in provincial revenues; therefore, he decided to initiate a policy of transfer of social services, among which was education. (Ministerio de Cultura y Educación 1980, 1, 151)

Despite the authoritarian regime, the governors voiced their concerns. Among others, the governor of Salta wrote to the minister of interior in November 1977: “by no means is the provincial treasury in a situation to afford the total costs of the services to be transferred” (Kisilevsky 1990, 20). At this time, however, the military’s grip on power was at its strongest, and the unfunded transfer was imposed from above. The administrative decentralization of 1978 had disastrous fiscal consequences for the provinces. The allocation of provincial resources for education had to increase from 14% in 1977 to almost 20% in 1982 (IMF 1985), at the same time that automatic transfers to the provinces decreased from 48.5% to 29% of all shared revenues (FIEL 1993, 151). Thirteen percent of the primary schools (about 3,400 schools) closed down prior to 1980, and governors were forced to beg for discretionary transfers from the national executive to avoid further closures.

Unfunded administrative decentralization had four important policy effects: (1) it *reshaped the preferences* of governors toward political and fiscal decentralization; (2) it contributed to the *reproduction of power* of the national executive; (3) it produced a *demonstration effect* by providing an example that future policymakers could follow; and (4) it produced *incrementalism* within the educational sector toward further decentralization of responsibilities.

During the electoral campaign of 1983, at least six political parties (including the two main parties: *Unión Cívica Radical*, UCR; and *Partido Justicialista*, PJ) advocated for a constitutional reform (Leiva and Abásalo 2000). The common concern was to strengthen political institutions and to avoid future disruptions to democratic rule. Several proposals to reform the constitution were introduced in congress in the first 2 years of the democratic transition. At the end of 1985, President Raúl Alfonsín (1983–1989) ordered the creation of a council to study the matter. The council’s recommendation for a constitutional reform included the creation of a mixed presidential system (with a prime minister), the strengthening of federalism, decentralization of the state, municipal autonomy, provincial control over natural resources, and limits on the president’s authority to intervene in the affairs of the provinces (Consejo para la Consolidación de la Democracia 1986). The council’s proposal was highly decentralizing, from both political and fiscal perspectives. Had it been implemented, it would have granted mayors constitutional autonomy, a prerogative they lack to this day. Governors would have had total control over natural resources (including oil) and more autonomy from the national executive in situations leading to federal interventions. Had this reform materialized, its political effects would have likely been similar to those of Brazil’s 1988 constitution (on such effects, see Stepan 2000). Interestingly, the debate over the constitutional reform in Argentina became structured along partisan (rather than territorial) lines (Botana and Mustapic 1991; Smulovitz 1987), and the governors did not endorse this political decentralization reform. Instead, with the return to democracy, governors focused on a fiscal reform, exhibiting a shift in their expected preferences.

Given the design of the prior round of administrative decentralization, governors were eager to negotiate an increase in fiscal transfers. When the revenue-sharing law of 1973 expired at the end of 1984, governors pushed to have a new revenue-sharing law in place. Carlos Menem, who at the time was the governor of La Rioja, proposed that the interior provinces rebel and cut the supply of energy to the city of Buenos Aires until an agreement on fiscal transfers was reached with the president (Pérez 1986, 68). But President Alfonsín controlled the timing of the reform and was successful in delaying its approval. Meanwhile, he used discretionary transfers to buy the political support of opposition governors. Discretionary transfers amounted to 59% of the total transfers in 1985 and 54% in 1986 (Ministerio de Economía 1989, 177–79). Thus, from 1984 to 1987, Alfonsín gained bargaining power vis-à-vis the governors by using the fiscal transfers to the provinces—which they desperately needed after unfunded administrative decentralization—in exchange for political support (mainly in the Senate).

Only after the 1987 midterm elections, when the ruling party lost its majority in the House (passing from 51% to 46% of the seats) and five governorships to the PJ, President Alfonsín agreed to the governors’ demand for redistribution of revenue-shared taxes. On January 7, 1988, congress passed a new revenue-sharing



law (*Ley de Coparticipación*, or Law 23,548) by which the provinces were granted 57.66% and the national government 42.34% of all revenue-shared taxes, and the discretionary transfers were cut to 1% of the shared taxes. By all accounts, this fiscal decentralization law was a victory for the governors, which came about when an exogenous change (the midterm elections of 1987) altered the balance of power between the president and the governors inherited from the first round of reforms. But the reform was also instrumental to the national executive. By that point, mounting economic problems and adverse midterm electoral results had made it clear that the ruling party would not retain the presidency after 1989. If the PJ were to win the 1989 presidential election, the new co-participation law would guarantee resources to UCR governors.<sup>13</sup>

The provincial fiscal recovery did not last long, however. Soon after the new revenue-sharing law was passed, the national executive (now in the hands of the PJ) was able to push forward a second round of unfunded administrative decentralization, which neutralized the effects of fiscal decentralization. On December 6, 1991, the Argentine congress passed Law 24,049 according to which the administration of all national secondary and adult schools and the supervision of private schools were transferred to the provinces and the city of Buenos Aires. Two food programs and the few remaining national hospitals were also transferred. The estimated cost of the transfer was 1.2 billion dollars per year, the equivalent of almost 10% of the total provincial expenditures and 15% of the total national transfers. Over 2,000 national schools, 72,000 teachers, and 700,000 students were incorporated into the provincial systems of education, which also had to supervise more than 2,500 private schools. Article 14 of the law established that the cost of the transferred services would be paid with provincial resources, whereas Article 15 stated that whenever the revenues collected in a given month were below the average of the April–December 1991 period, the national government would transfer 1.2 billion pesos or the difference required to match that amount. Government documents and interviews with national and subnational officials suggest that such guarantee was not enacted and the transfer of responsibilities was largely unfunded (see Falletti 2003, 136–55).

The first round of administrative decentralization of 1978 had a demonstration effect for the second round of administrative decentralization. In 1991, as a result of the convertibility law, the absolute amount of revenues in the provinces had doubled—the automatic transfers passed from 4,810 million dollars in 1990 to 8,846 million in 1992 (Subsecretaría de Relaciones Fiscales y Económicas con las Provincias 1994, 15). In this context, as in 1978, it was easier to pass an unfunded administrative decentralization reform. Minister of Economy Domingo Cavallo appealed to the same arguments used in 1978 by Minister of Economy Martínez de Hoz to justify the transfer of responsibilities. In meetings with the governors, Cavallo argued

that the increase in revenues would allow the provinces to afford the expenditures generated by the transfer of social services.

Finally, the 1978 decentralization reform also produced incrementalism. Although the national secondary schools were administered *de jure* by the national government until 1992, a process of decentralization of responsibilities was already under way. In the words of the governor of Mendoza from 1987 to 1991:

... the truth is that a *de facto* transfer [of national schools] was already taking place, without recognition in the distribution of revenues. In practice ... every time there was a problem in a national school, [people] came to the provincial government to ask for a solution. (Bordón, José Octavio, interview by author, Buenos Aires, February 8, 2001)

National officials also recognized this situation. Secretary of education Luis A. Barry said:

There were [national] schools that for ten years had not had any supervision. They were managed by phone [from Buenos Aires] or ... by mail. The link was formal, epistolary, but not efficient. (X National Seminar on National Budget, Buenos Aires, Public Administrators Association)

Or as a member of the ministry of economy put it: “only in their plates were the schools national” (Pezoa, Juan Carlos, interview by author, Buenos Aires, February 13, 2001). Under these conditions, the governors were more inclined to accept a transfer of schools, even if it was to be funded primarily with provincial resources. The 1978 round of administrative decentralization enabled the national executive to pass a similar policy reform, albeit in a democratic context, 13 years later.

Political decentralization came last in the first cycle of market-oriented decentralization reforms in Argentina. It occurred in 1994, when President Menem (1989–1995 and 1995–1999) exchanged constitutional reforms as a bargaining chip for his reelection. Political autonomy was granted to the city of Buenos Aires, but various decentralization reforms proposed in the constitutional assembly by provincial representatives (and also included in the 1986 report of the Council for Democratic Consolidation) failed to pass. Reforms such as a higher share of subnational revenues, provincial control of natural resources, and constitutionally guaranteed municipal autonomy were all proposed in the constituent assembly; but due to the political pressure of the national executive all these fiscal and political decentralization proposals did not pass. In other words, the national executive was able to control the timing as well as the main contents of the political decentralization reform of 1994.

In sum, as a consequence of the first round of administrative decentralization, the preferences of the governors were reshaped. Because the 1978 transfer of schools was unfunded, governors were more concerned, after the return to democracy, with revenues than with a constitutional reform that would have granted them more political autonomy (e.g., by

<sup>13</sup> I thank one of the anonymous reviewers for this interpretation.

protecting them against federal interventions or granting them control of natural resources). Arguably, governors could have pursued both types of reforms at the same time, but they did not. Instead, between 1984 and 1987, they focused on fiscal decentralization and did not endorse the project of political decentralization. The 1978 reform also had demonstration and incremental effects in that additional unfunded administrative decentralization measures were made possible. Finally, the first round of administrative decentralization initiated a reproduction of the bargaining power of presidents, who were then able to control not only the timing of fiscal and political decentralization but also the contents and extent of these reforms.

The sequence of administrative, fiscal, and political reforms followed by Argentina resulted in a small change in the relative power of the governors and mayors. The share of expenditures increased, but by a lower amount than the changes experienced by Colombia, Mexico, or Brazil. The subnational share of revenues decreased. This was in spite of the fact that, from 1978, the Argentine provinces were allocated responsibilities whose cost amounted to approximately 35% of the total transfers they received from the center. Regarding policy-making authority in the educational sector, it remained unchanged until 1993, when the new federal law of education was passed. The appointment of subnational officials remained the same with the exception of the mayor of Buenos Aires, who became popularly elected in 1996. Finally, the territorial representation of interests in congress stayed more or less constant throughout the period. Despite the introduction of decentralization policies that transferred responsibilities, resources, and authority to subnational governments, the sequence in which the reforms took place meant that the intergovernmental balance of power remained unchanged in Argentina. Compared to their situation prior to 1976, governors had acquired more responsibilities and fewer fiscal resources, with no change in their political authority.

## CONCLUSION

Decentralization policies have the potential to reverse long-standing, deeply embedded features of intergovernmental relations. In a relatively short time span, reforms such as the direct election of governors and mayors, the transfer of national schools to states and municipalities, or the devolution of fiscal authority to the subnational units can undo the "skillful organization of authority" or the "complicated administrative machine" once described by Alexis de Tocqueville (in Schleifer 1980, 137–38). However, the impact of these reforms on the power of governors and mayors is not always the same.

The first conclusion drawn from this article is that decentralization does not always transfer power to governors and mayors. The unpacking of the concept of decentralization in its administrative, fiscal, and political dimensions reveals that certain types of reforms decrease the power of subnational officials. Policies such as unfunded administrative decentralization make

subnational executives more dependent on the national government for fiscal resources. The three-dimensional definition advanced in this article also allows one to distinguish between the interests of national and subnational executives regarding types of decentralization.

The second conclusion is that the degree of change in intergovernmental balance of power is largely dependent on the sequence in which administrative, fiscal, and political decentralization reforms take place. I have shown that if subnational interests prevail in the first round of reforms, political decentralization is likely to occur first. This first reform enhances the power and capacities of subnational politicians and public officials for the negotiations over the next rounds of reforms. The devolution of political power early in the sequence is likely to produce coordination among the beneficiaries of this policy who will push forward in the direction of further decentralization. As O'Neill writes: "the most formidable obstacle to recentralization comes from the newly enfranchised; once passed, [political] decentralization builds a constituency for itself, making it difficult—but not impossible—to reverse within a democracy" (2003, 1076). Thus, according to the preferences of subnational actors, fiscal and administrative decentralization are likely to follow in that order. This sequence of decentralization that devolves political autonomy first, fiscal resources next, and administrative responsibilities third, is likely to produce a significant change in the degree of autonomy of subnational officials—as the Colombian case has illustrated.

In contrast, if national interests prevail at the beginning of the process, administrative decentralization is likely to occur first. If, through administrative decentralization, the center is able to offload responsibilities without transferring the fiscal resources to meet those responsibilities, the central government strengthens its dominance over subnational governments for the next rounds of reforms. The devolution of responsibilities at the beginning of the sequence is likely to set constraints on what subnational officials are politically capable of doing and fiscally able to afford. Under fiscal strain, subnational governments are more likely to agree to the terms set by the central level when fiscal decentralization follows. In this situation, the national executive also prevails in setting the terms for the final round of political reforms, if they were to happen. The outcome of this sequence is likely to be a low degree of change in the autonomy of subnational officials, despite the implementation of the reforms—as the case of Argentina has shown. Moreover, because in this type of sequence decentralization does not create a constituency for itself, reversals (or recentralization) seem more likely to occur in this type of cases than when political decentralization takes place at the beginning of the process.

Once we unpack the process of decentralization into its component policies, examine carefully the territorial preferences of national and subnational politicians toward different types of decentralization, and analyze the effects of each policy on the intergovernmental balance of power and subsequent rounds of reforms, we find that decentralization processes conform to

path-dependent sequences. Like in other path-dependent processes, "earlier events matter much more than later ones" (Pierson 2000, 253), or "when things happen within a sequence affects *how* they happen" (Tilly 1984, 14). I have shown how two opposing decentralization sequences unfolded in two Latin American countries. I have also identified the self-reinforcing mechanisms (incrementalism, coordination, reshaping of preferences, reproduction of power, and demonstration effects) through which these two sequences brought about the intergovernmental balance of power outcomes expected according to the sequential theory of decentralization.

However, there are areas where more research is necessary. First, it is necessary to confirm whether the other four sequences of decentralization presented in this article lead to the expected results. Catherine Hirbour (2003) applied this framework to the case of Peru. She found that although the movement toward decentralization was initiated from below and political decentralization took place first, reactive mechanisms led to the predominance of the national level in the second and third rounds of reforms. A sequence of political, administrative, and fiscal decentralization, taking place in that order, led to a low degree of change in the intergovernmental balance of power, consistent with the theoretical expectation. I also expect analyses of the processes of decentralization in Mexico and Brazil to show that these countries have followed sequences that lead to medium or high level of degrees of change in intergovernmental balance of power. Previous works point in this direction (Falleti 2003; Montero 2001; Samuels 2004), but further in-depth comparative research is needed.

Second, national and subnational actors have different preferences not only with regard to the type of decentralization (which was analyzed here) but also with regard to the *level* of government targeted by decentralization (i.e., intermediate versus local levels). If presidents have to choose between decentralization to the state and decentralization to the local level, they will probably choose decentralization toward the municipal level. This is because mayors pose less of an electoral and financial threat to presidents than governors do. Governors and mayors, on the other hand, will prefer decentralization toward their own levels of government. These preferences may affect the composition of the coalitions behind decentralization policies, as presidents may choose to ally with mayors against governors. Future research should elucidate the political circumstances under which this is likely to happen and what the consequences of such coalitions are.

Third, I have focused on the first cycle of post-developmental decentralization reforms, which ends once the three types of decentralization (administrative, fiscal, and political) have all occurred. Nonetheless, further decentralization and centralization reforms are likely to occur after the first cycle of reforms. The importance of the first cycle of decentralization is that it sets the tone for what is likely to follow. For example, both Argentina and Brazil have recently undergone re-centralization reforms (Eaton and Dickovick

2004), but in Brazil the negotiations incorporated the governors' and mayors' proposals to a larger extent than they did in Argentina. Future research will have to specify the degree to which the consequences of the first cycle of decentralization constrain future rounds of reforms and the degree to which exogenous political and economic changes could contribute to relax those constraints.

A final word is merited on the applicability of the sequential theory of decentralization to other cases and areas of study. I have focused on the bargaining between national officials on the one hand and subnational officials (both of the intermediate and local levels of government) on the other. Increasingly, however, local or municipal governments are the focus of policy reforms and are being granted larger amounts of resources and responsibilities. The preferences of bargaining actors and the sequential logic presented here could prove useful in analyzing negotiations between governors and mayors. This would allow us to account for within-country differences in the level of power devolved from state to local governments. Finally, can the sequential theory of decentralization be applied to other countries and regions of the world? The domain of this theory are those countries that have at least two levels of government (even if the subnational level is not politically autonomous from the central level) and have seen at least two types of decentralization reforms occur at different points in time. In such cases, we should expect the type of interests that prevail in the first round of decentralization and the sequence of policy reforms that follows to be the main determinants of the resulting degree of change in the intergovernmental balance of power.

## REFERENCES

- Angell, Alan, Pamela Lowden, and Rosemary Thorp. 2001. *Decentralizing Development*. Oxford and New York: Oxford University Press.
- Ansaldi, Waldo. 1992. "Frivolous and Empty-headed, Iron Hand in Silky Glove. A Proposal to Conceptualize the Term Oligarchy in Latin America" [in Spanish]. In *América Latina. Planteos, Problemas, Preguntas*, ed. P. Funes. Buenos Aires: Manuel Suárez, 13–20.
- Artana, Daniel, et al. 1995. "Argentina." In *Fiscal Decentralization in Latin America*, ed. R. López Murphy. Washington, DC: Inter-American Development Bank, 59–136.
- Block, Fred. 1994. "The Roles of the State in the Economy." In *Handbook of Economic Sociology*, ed. N. Smelser and R. Swedberg. Princeton: Princeton University Press, 691–710.
- Botana, Natalio, and Ana María Mustapic. 1991. "Constitutional Reform and Political Regime in Argentina" [in Spanish]. In *Reforma institucional y cambio político*, ed. D. Nohlen and L. De Riz. Buenos Aires: CEDES, Legasa, 45–92.
- Burki, Shahid J., Guillermo E. Perry, and William R. Dillinger. 1999. *Beyond the Center: Decentralizing the State*. Washington, DC: The World Bank.
- Camdessus, Michel. 1999. "Second Generation Reforms: Reflections and New Challenges." Presented at IMF Conference on Second Generation Reforms, November 8, at Washington, DC.
- Cheema, Shabbir G., and Dennis A. Rondinelli, eds. 1983. *Decentralization and development: policy implementation in developing countries*. Beverly Hills: Sage.
- Collier, Ruth Berins, and David Collier. 1991. *Shaping the Political Arena. Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton, NJ: Princeton University Press.

- Collins, Charles David. 1988. "Local government and Urban protest in Colombia." *Public Administration and Development* 8 (4): 421–36.
- Consejo para la Consolidación de la Democracia. 1986. *Constitutional Reform: Preliminary Report of the Council for Democratic Consolidation* [in Spanish]. Buenos Aires: EUDEBA.
- Cornelius, Wayne A., Todd A. Eisenstadt, and Jane Hindley, eds. 1999. *Subnational Politics and Democratization in Mexico*. La Jolla, CA: Center for U.S.-Mexican Studies, UCSD.
- Corrales, Javier. 2004. "Multiple Preferences, Variable Strengths: The Politics of Education Reform in Argentina." In *Crucial Needs, Weak Incentives. Social Sector Reform, Democratization, and Globalization in Latin America*, ed. J. Nelson and R. Kaufman. Washington, DC and Baltimore, MD: Woodrow Wilson Center Press and Johns Hopkins University Press, 315–49.
- Correa Henao, Néstor Raúl. 1994. "Decentralization and Public Order" [in Spanish]. In *Diez años de descentralización. Resultados y Perspectivas*, ed. A. Noyes et al. Bogotá: FESCOL, 9–61.
- Dahl, Robert. 1986. "Federalism and the Democratic Process" In *Democracy, Liberty, and Equality*. Oslo, Denmark: Norwegian University Press, 114–26.
- Di Gropello, Emanuela, and Rosella Cominetti, eds. 1998. *Decentralization of Education and Health: A Comparative Analysis of Latin American Experiences* [in Spanish]. Santiago de Chile: UNECLAC.
- Diamond, Larry, and Svetlana Tsalik. 1999. "Size and Democracy. The Case for Decentralization." In *Developing Democracy Toward Consolidation*, ed. L. Diamond. Baltimore and London: The Johns Hopkins University Press, 117–60.
- Dillinger, William R., and Steven Webb. 1999. "Decentralization and Fiscal Management in Colombia." Washington, DC: World Bank.
- DNP and PNUD. 1998. *Human Development Report for Colombia* [in Spanish]. Santafé de Bogotá: Departamento Nacional de Planeación, Misión Social y PNUD.
- Eaton, Kent. 2001. "Decentralisation, Democratisation and Liberalisation: The History of Revenue Sharing in Argentina, 1934–1999." *Journal of Latin American Studies* 33 (1): 1–28.
- . 2002. "Fiscal Policy Making in the Argentine Legislature." In *Legislative Politics in Latin America*, ed. S. Morgenstern and B. Nacif. Cambridge, UK, New York, NY: Cambridge University Press, 287–314.
- . 2004. *Politics Beyond the Capital: The Design of Subnational Institutions in South America*. Stanford, CA: Stanford University Press.
- Eaton, Kent, and J. Tyler Dickovick. 2004. "The Politics of Recentralization in Argentina and Brazil." *Latin American Research Review* 39 (1): 90–122.
- Escobar-Lemmon, Maria. 2001. "Fiscal Decentralization and Federalism in Latin America." *Publius. The Journal of Federalism* 31 (4): 23–41.
- . 2003. "Political Support for Decentralization: An Analysis of the Colombian and Venezuelan Legislatures." *American Journal of Political Science* 47 (4): 683–97.
- FCM. 2004. *Mandate of the Federación of Colombian Municipalities 1991* [in Spanish]. Cited 11/8 2004. Available at <http://www.fcm.org.co/es/load.php?uid=0/leng=es/0/Misionvision.htm>
- Falleti, Tulia G. 2003. "Governing Governors: Coalitions and Sequences of Decentralization in Argentina, Colombia, and Mexico." Ph.D. diss. Northwestern University.
- FIEL, Fundación de Investigaciones Económicas Latinoamericanas. 1993. *Toward a New Organization of Fiscal Federalism in Argentina* [in Spanish]. Buenos Aires: FIEL.
- Garbán Pavía, Pilar, and Carlos Moreno Ospina. 1992. *Local Power: The Reality and Utopia of Decentralization in Colombia* [in Spanish]. Bogotá: Tercer Mundo Editores—Instituto de Estudios Políticos y Relaciones Internacionales, Universidad Nacional de Colombia.
- Garman, Christopher, Stephan Haggard, and Eliza Willis. 2001. "Fiscal Decentralization: A Political Theory with Latin American Cases." *World Politics* 53 (January): 205–36.
- Gibson, Edward L., ed. 2004. *Federalism and Democracy in Latin America*. Baltimore and London: The Johns Hopkins University Press.
- Gibson, Edward L., and Ernesto Calvo. 2000. "Federalism and low-maintenance constituencies: Territorial dimensions of economic reform in Argentina." *Studies in Comparative International Development* 35 (3): 32–55.
- Gibson, Edward L., and Tulia Falleti. 2004. "Unity by the Stick: Regional Conflict and the Origins of Argentine Federalism." In *Federalism and Democracy in Latin America*, ed. E. Gibson. Baltimore and London: The Johns Hopkins University Press, 226–54.
- González, Fernán E. 1990. "The End of the DESARROLLISTA State: State and Capital Accumulation in Colombia, 1958–84." In *The State and Capital Accumulation in Latin America Volume 2*, eds. C. Anglade and C. Fortin. Pittsburgh, PA: University of Pittsburgh Press, 54–87.
- Grindle, Merilee. 2000. *Audacious Reforms: Institutional Invention and Democracy in Latin America*. Baltimore and London: The Johns Hopkins University Press.
- Haggard, Stephan. 1998. "The Reform of the State in Latin America" [in Spanish]. *Reforma y Democracia* 11: 189–230.
- Hirbourn, Catherine. 2003. "Sequences of Reform and the Dynamic Decentralization Process in Peru." Masters thesis. University of British Columbia.
- Huber, Evelyn, and John D. Stephens. 2001. *Development and Crisis of the Welfare State. Parties and Policies in Global Markets*. Chicago and London: University of Chicago Press.
- IMF, International Monetary Fund. Various Years. *Government Finance Statistics Yearbook*. Washington, DC: IMF.
- INDEC. 1997. *Statistics Yearbook. 1997* [in Spanish]. Buenos Aires, Argentina: Instituto Nacional de Estadísticas y Censo, Ministerio de Economía y Obras y Servicios Públicos.
- INEGI. 1995. *Statistics Yearbook of the United States of Mexico* [in Spanish]. Mexico: Instituto Nacional de Estadística, Geografía e Informática.
- Jones, Mark, et al. 2002. "Amateur Legislators-Professional Politicians: The Consequences of Party-Centered Electoral Rules in a Federal System." *American Journal of Political Science* 46 (3): 656–69.
- Kisilevsky, Marta. 1990. *Nation and Provinces since the 1978 Transfer of Educational Services* [in Spanish]. Buenos Aires: Consejo Federal de Inversiones.
- . 1998. "Federalism and Education: a historical space for distributive conflicts" [in Spanish]. Masters thesis. Facultad Latinoamericana de Ciencias Sociales.
- Leiva, Alberto David, and Ezequiel Abásalo. 2000. *Argentine Constitutionalism in the Twentieth Century* [in Spanish]. Buenos Aires: Ed. Dunker.
- Lujambio, Alonso. 2000. *Shared Power. An Essay on Mexico's Democratization* [in Spanish]. Mexico: Editorial Oceano de Mexico.
- Mahoney, James. 2000. "Path dependence in historical sociology." *Theory and Society* 29 (4): 507–48.
- Maldonado, Alberto. 2000. *The Effects of Political Decentralization in Colombia* [in Spanish]. Bogotá: Departamento de Investigaciones Fundación Universidad Central.
- Manor, James. 1999. *The Political Economy of Democratic Decentralization*. Washington, DC: The World Bank.
- Ministerio de Cultura y Educación. 1980. *Evaluation of the Transfer* [in Spanish]. Buenos Aires: Secretaría de Estado de Educación, Dirección Nacional de Investigación, Experimentación y Perfeccionamiento Educativo.
- Ministerio de Economía. 1989. *Politics for Structural Change in the Public Sector: Speeches on National Budget Proposals 1986–1989* [in Spanish]. Buenos Aires: Poder Ejecutivo Nacional, Ministerio de Economía, Secretaría de Hacienda.
- Misión de Finanzas Intergubernamentales. 1981. *Intergovernmental Finances in Colombia* [in Spanish]. Bogotá: DNP, Departamento Nacional de Planeación.
- Montero, Alfred P. 2001. "After Decentralization: Patterns of Intergovernmental Conflict in Argentina, Brazil, Spain, and Mexico." *Publius. The Journal of Federalism* 31 (4): 43–64.
- Montero, Alfred P., and David J. Samuels. 2004. "The Political Determinants of Decentralization in Latin America: Causes and Consequences." In *Decentralization and Democracy in Latin America*, ed. A. P. Montero and D. J. Samuels. Notre Dame, IN: University of Notre Dame Press, 3–32.
- Nickson, R. Andrew. 1995. *Local Government in Latin America*. Boulder and London: Lynne Rienner.

- Novick de Senén González, Silvia. 1995 "Argentina." In *Decentralization and Planning of Education: Recent Experiences in Five Latin American Countries*, ed. C. N. Maplica Faustor. Paris: UNESCO: International Institute for Educational Planning, 121–168.
- Oates, Wallace. 1972. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.
- Ocampo Gaviria, José Antonio, and Guillermo Perry Rubio. 1983 "The Fiscal Reform, 1982–1983" [in Spanish] *Coyuntura Económica*. 215–64.
- O'Neill, Kathleen. 1999. "Decentralization in the Andes: Power to the People or Party Politics?" Ph.D. diss. Government, Harvard University.
- . 2003. "Decentralization as an Electoral Strategy." *Comparative Political Studies* 36 (9): 1068–91.
- Penfold-Becerra, Michael. 1999. "Institutional Electoral Incentives and Decentralization Outcomes: Comparing Colombia and Venezuela." Ph.D. diss. Political Science, Columbia University.
- Pierson, Paul. 1992 "Policy Feedbacks and Political Change: Contrasting Reagan and Thatcher's Pension-Reform Initiatives" *Studies in American Political Development* 6 (Fall): 359–90.
- . 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94 (2): 251–67.
- . 2004. *Politics in Time: History, Institutions, and Social Analysis*. Princeton and Oxford: Princeton University Press.
- Pírez, Pedro. 1986. *Federal Coparticipation and Decentralization of the State* [in Spanish]. Buenos Aires: Centro Editor de América Latina.
- Registraduría Nacional del Estado Civil. Various Years. *Electoral Statistics* [in Spanish]. Bogotá, Colombia.
- Riker, William H. 1964. *Federalism. Origin, Operation, Significance*. 3rd ed, Boston and Toronto. Little, Brown and Co.
- Rodden, Jonathan. 2002. "The Dilemma of Fiscal Federalism: Grants and Fiscal Performance around the World." *American Journal of Political Science* 46 (3): 670–87.
- Samuels, David. 2003. *Ambition, Federalism, and Legislative Politics in Brazil*. New York: Cambridge University Press.
- Samuels, David J. 2004. "The Political Logic of Decentralization in Brazil." In *Decentralization and Democracy in Latin America*, ed. A. P. Montero and D. J. Samuels. Notre Dame, IN: University of Notre Dame Press, 67–93.
- Samuels, David, and Scott Mainwaring. 2004. "Strong Federalism, Constraints on the Central Government, and Economic Reform in Brazil." In *Federalism and Democracy in Latin America*, ed. E. Gibson. Baltimore: Johns Hopkins University Press, 85–130.
- Schleifer, James. 1980. *The making of Tocqueville's "Democracy in America"*. Chapel Hill: The University of North Carolina Press.
- Schneider, Ben Ross. 1999. "The *Desarrollista* State in Brazil and Mexico." In *The Developmental State*, ed. M. Woo-Cumings. Ithaca and London: Cornell University Press, 276–305.
- Skowronek, Stephen. 1993. *The Politics Presidents Make. Leadership from John Adams to George Bush*. Cambridge, MA and London: The Belknap Press of Harvard University Press.
- Smulovitz, Catalina. 1987. "Constitutional Reform and Democratic Consolidation in Argentina. A new game or back to the instability of rules?" [in Spanish]. *Cuadernos del CLAEH* 43 (2da serie, Año 12): 69–76.
- Snyder, Richard. 2001. "Scaling Down. The Subnational Comparative Method." *Studies in Comparative International Development* 36 (1): 93–110.
- Stepan, Alfred. 2000. "Brazil's Decentralized Federalism: Bringing Government Closer to the Citizens?" *Daedalus* 129 (2): 145–69.
- . 2004. "Toward a New Comparative Politics of Federalism, Multinationalism and Democracy: Beyond Rikerian Federalism." In *Federalism and Democracy in Latin America*, ed. E. L. Gibson. Baltimore and London: The John Hopkins University Press, 29–84.
- Stinchcombe, Arthur L. 1968. *Constructing Social Theories*. New York, Chicago, San Francisco, Atlanta: Harcourt, Brace & World, Inc.
- Subsecretaría de Relaciones Fiscales y Económicas con las Provincias. 1994. *Structural Changes in the relation between Nation and Provinces* [in Spanish]. Buenos Aires: Secretaría de Hacienda. Ministerio de Economía y Obras y Servicios Públicos.
- Tarrow, Sidney. 1978. "Introduction." In *Territorial Politics in Industrial Nations*, eds. S. Tarrow, P. J. Katzenstein and L. Graziano. New York and London: Praeger Publishers, 1–27.
- Tavares de Almeida, Maria Herminia. 1995. *Federalism and Social Policies* [in Spanish]. Santiago de Chile: Naciones Unidas CEPAL.
- Thelen, Kathleen. 2000. "Timing and Temporality in the Analysis of Institutional Evolution and Change." *Studies in American Political Development* 14 (Spring): 101–8.
- . 2003. "How Institutions Evolve: Insights from Comparative-Historical Analysis." In *Comparative Historical Analysis in the Social Sciences*, ed. J. Mahoney and D. Rueschemeyer. New York: Cambridge University Press, 208–40.
- Tilly, Charles. 1984. *Big Structures, Large Processes, Huge Comparisons*. New York: Russell Sage Foundation.
- Treisman, Daniel. 1999. *After the Deluge: Regional Crises and Political Consolidation in Russia*. Ann Arbor, MI: University of Michigan Press.
- . 2000. "Decentralization and Inflation: Commitment, Collective Action, or Continuity?" *American Political Science Review* 94 (4): 837–57.
- Vargas González, Jorge Enrique, and Alfredo Sarmiento Gómez. 1997. *Characteristics of Colombia's Decentralization* [in Spanish]. Santiago de Chile: Naciones Unidas. Comisión Económica para América Latina y el Caribe.
- Velásquez, Fabio. 1995. "Decentralization in Colombia: antecedents, developments, and perspectives" [in Spanish]. In *Descentralizar en América Latina? Gestión Urbana, Vol 3*, ed. J. Werner Hass and A. Rosenfeld. Quito, Ecuador. GTZ/PGU—Sociedad Alemana de Cooperación Técnica/Programa de Gestión Urbana, 237–310.
- Weyland, Kurt. 2002. *The Politics of Market Reforms in Fragile Democracies: Argentina, Brazil, Peru, and Venezuela*. Princeton and Oxford: Princeton University Press.
- Wibbels, Erik. 2000. "Federalism and the Politics of Macroeconomic Policy and Performance." *American Journal of Political Science* 44 (4): 687–702.
- . 2004. "Decentralization, Democracy, and Market Reform: On the Difficulties of Killing Two Birds with One Stone." In *Decentralization and Democracy in Latin America*, ed. A. P. Montero and D. J. Samuels. Notre Dame, IN: University of Notre Dame Press, 203–34.
- Wiesner Durán, Eduardo. 1992. *Colombia. Decentralization and Fiscal Federalism. Final Report of the Decentralization Team* [in Spanish]. Bogotá: Departamento Nacional de Planeación.
- Willis, Eliza, Christopher Garman, and Stephan Haggard. 1999. "The politics of decentralization in Latin America." *Latin American Research Review* 34 (1): 7–56.
- Wright, Deil S. 1978. *Understanding Intergovernmental Relations*. North Scituate, MA: Duxbury Press.



# Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success

BONNIE M. MEGUID *University of Rochester*

**W**hat accounts for variation in the electoral success of niche parties? Although institutional and sociological explanations of single-issue party strength have been dominant, they tend to remove parties from the analysis. In this article, I argue that the behavior of mainstream parties influences the electoral fortunes of the new, niche party actors. In contrast to standard spatial theories, my theory recognizes that party tactics work by altering the salience and ownership of issues for political competition, not just party issue positions. It follows that niche party support can be shaped by both proximal and non-proximal competitors. Analysis of green and radical right party vote in 17 Western European countries from 1970 to 2000 confirms that mainstream party strategies matter; the modified spatial theory accounts for the failure and success of niche parties across countries and over time better than institutional, sociological, and even standard spatial explanations.

Since the 1960s, political systems around the world have undergone a revolution. From Western Europe and North America to Australasia and Latin America, new political parties have emerged and gained popularity on the basis of previously overlooked issues such as the environment, immigration, and regional autonomy. In addition to challenging the economic focus of the political debate, these *niche parties* have threatened the electoral and governmental dominance of mainstream political parties. For example, since 1960, over 54% of green, radical right and ethnoterritorial parties in Western Europe have held a seat in a national legislature.<sup>1</sup> Almost 10% have participated in coalition governments, and the participation of over half of those parties was pivotal to the formation of majority governments. Even when niche parties have failed to attain many or any seats, their electoral strength has influenced the fortunes of others. The role of the French radical right party, the Front National, in the legislative victory of the Socialist Party (and the defeat of the Gaullists) in 1997 is just one of many similar cases. Given the weighty implications of new party electoral support, this article examines why some parties flourish while others flounder. In other words, what determines variation in the electoral success of niche parties?

This question has typically been answered with institutional or sociological explanations. According to the first set of theories, electoral rules, governmental types, and the structure of the state, among other institutions, constrain or facilitate a new party's electoral advancement (e.g., Duverger 1963; Harmel and Robertson 1985; Müller-Rommel 1996). For proponents of sociological approaches, new party support

varies by the socioeconomic conditions and value orientation of a society (Golder 2003; Inglehart 1998). Although popular, these explanations are insufficient. Static institutions cannot account for variation in a party's vote share over time. And, as shown in cross-national analyses of new party vote (Swank and Betz 1995, 1996, 2003), both sociological and institutional approaches stumble in the face of the numerous green and radical right parties that attract little support under propitious circumstances and significant support under inauspicious ones.

In emphasizing the context in which party competition takes place, the existing literature has curiously ignored the behavior of the competitors. This article brings parties back into party analysis. I demonstrate the critical role that the most powerful set of party actors—mainstream parties of the center-left and center-right—plays in shaping the success of niche parties.

## THE NEW COMPETITORS: THE NICHE PARTY PHENOMENON

The electoral arenas of developed and developing democracies have been flooded with new political parties over the past 40 years. Although many of these new political organizations are variants of the existing socialist, liberal, and conservative parties, there is a group of parties that stands out. These actors, which I call *niche parties*, differ from their fellow neophytes and the mainstream parties in three significant ways.<sup>2</sup> First, niche parties reject the traditional class-based orientation of politics. Instead of prioritizing economic demands, these parties politicize sets of issues which were previously outside the dimensions of party competition. Green parties, for example, emerged in the 1970s to call attention to the underdiscussed issues of environmental protection, nuclear disarmament, and

Bonnie M. Meguid is Assistant Professor, Department of Political Science, University of Rochester, Rochester, NY 14627. (bonnie.meguid@rochester.edu).

Previous versions of this paper were presented at the 2002 Annual Meeting of the Midwest Political Science Association and the 2001 Annual Meeting of the American Political Science Association. I thank conference panelists, Jim Adams, Kevin Clarke, Michael Delgado, Jorge Domínguez, Matt Golder, Peter Hall, Torben Iversen, Ken Kollman, and the Editor and reviewers of the *APSR* for their helpful comments and suggestions.

<sup>1</sup> Author's calculations from Woldendorp, Keman, and Budge 1998

<sup>2</sup> Following the tradition of identifying parties by their substantive positions, scholars have typically treated green and radical right parties as distinct party families (see Kitschelt 1994, 1995, O'Neill 1997). When we focus on the function of these new parties within the party system, however, their similarities outweigh their differences.

nuclear power. Radical right parties followed on their heels in the 1980s and 1990s, demanding the protection of (patriarchal) family values and a nationally oriented, immigrant-free way of life. Despite differences in the substantive nature of their demands, these parties similarly challenge the content of political debate.

Second, the issues raised by the niche parties are not only novel, but they often do not coincide with existing lines of political division. Niche parties appeal to groups of voters that may cross-cut traditional partisan alignments. As a result, cases of voter defection between “unlikely” party pairs have occurred. The defection of former British Conservative voters to the Green Party in 1989 and former French Communist party voters to the radical right Front National in 1986 are typical examples.

Third, niche parties further differentiate themselves by limiting their issue appeals. They eschew the comprehensive policy platforms common to their mainstream party peers, instead adopting positions only on a restricted set of issues. Even as the number of issues covered in their manifestos has increased over the parties’ lifetimes, they have still been perceived as single-issue parties by the voters. Unable to benefit from pre-existing partisan allegiances or the broad allure of comprehensive ideological positions, niche parties rely on the salience and attractiveness of their *one* policy stance for voter support.

The niche party phenomenon has most strongly affected the political arenas of Western Europe. Over the past 30 years, approximately 110 niche parties have contested elections in 18 countries.<sup>3</sup> Environmental and radical right parties are the most common types. Although the phenomenon is widespread, the number of parties competing in national-level elections has varied from a single example in Ireland to 20 in Italy (Mackie and Rose 1991, 1997). Niche party electoral success has also varied, with only 24% achieving a peak national vote of over 5%. It is important to note that this success is not concentrated in a few countries; thirteen countries have had at least one niche party surpass the 5% threshold, and all 18 have had at least one niche party office holder.<sup>4</sup>

## A STRATEGIC EXPLANATION OF NICHE PARTY VOTE

Recognition of these differences in niche party formation and success prompts an obvious question: why did some of these new parties gain more electoral support than others? Moreover, what determined the timing of the peaks and troughs in the electoral trajectories of these parties? In recent years, the standard answers to any question of new political party success have

been institutional and sociological (e.g., Golder 2003; Inglehart 1998). As noted in the introduction, however, the utility of structural explanations is limited. Not only do these theories fail to account for the electoral performance of several key cases, but they also downplay the role of political actors. When modeling the behavior of voters and its impact on party electoral prospects, the existing literature disregards the fact that parties have tools that allow them to adapt to the institutional and sociological environment in which they participate. In this article, I advance a theory of niche party success that focuses on the role of mainstream party strategies in determining the competitiveness of new political dimensions and that of the niche parties competing on them.<sup>5</sup>

Largely ignored by the literature on new party success,<sup>6</sup> strategic models of party competition are hardly new. Made famous by Downs (1957), the spatial theory of party and voter behavior—whereby rational parties choose policy positions to minimize the distance between themselves and the voters—lies at the heart of significant theoretical work on the entrance, interaction, and success of (mainstream) parties (e.g., Enelow and Hinich 1984; Kitschelt 1994; Shepsle 1991). According to this framework, parties competing for votes are faced with two possible strategies: movement toward (policy convergence) or movement away from (policy divergence) a specific competitor in a given policy space. Policy convergence, or what I call an *accommodative strategy*, is typically employed by parties hoping to draw voters away from a threatening competitor. On the other hand, by increasing the policy distance between parties, policy divergence, or what I term an *adversarial strategy*, encourages voter flight to the competing party.

This programmatic conception of party behavior has become the dominant lens through which to understand political competition and party strategies. However, it is not without limitations. Whether spatial theorists view the policy arena as having equally or unequally weighted dimensions, they explicitly assume that the salience of those issue axes remains fixed during party interaction. But just as exogenous factors like economic crises or natural disasters can alter the importance of an issue dimension, studies have shown that parties also can manipulate the perceived salience of issues within the political arena (Budge, Robertson, and Hearl 1987; Rabinowitz and Macdonald 1989). Budge, Robertson, and Hearl (39) observe that “(p)arties compete by accentuating issues on which they have an undoubted advantage, rather than by putting forward contrasting policies on the same issues.” In other words,

<sup>3</sup> These countries are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom

<sup>4</sup> The five countries lacking a niche party with a peak national vote greater than 5% are Greece, Ireland, the Netherlands, Portugal, and the United Kingdom.

<sup>5</sup> Mainstream parties are defined as the electorally dominant actors in the center-left, center, and center-right blocs on the Left-Right political spectrum. In this classification, the center-left parties explicitly exclude left-libertarian parties, whereas the center-right categorization excludes right-authoritarian, or right-wing, populist parties. The criteria generally yield three mainstream parties per country, one in each category. For more on coding, see the independent variables section.

<sup>6</sup> Notable exceptions include Rohrschneider 1993 and Harmel and Svasand 1997.

parties do not compete on all issues in the political space in every election.<sup>7</sup> By choosing which issues to compete on in a given election, parties can shape the importance of policy dimensions. Because voters, who often take their cues from political parties, discount the attractiveness of policies on issues they find irrelevant, a party's ability to downplay or highlight issues influences party fortunes.

In addition, existing strategic models have generally disregarded issue ownership. According to standard spatial theories, where voter decisions depend solely on ideological proximity, voters facing equally distant parties are indifferent between their political options. But voter choice is not necessarily (or typically) dictated by the flip of a coin. Just as partisan identification has been shown to influence voter decision-making in highly aligned political environments, a party's issue credibility, or ownership, plays a key role in issue-based voting (Budge and Farlie 1983); voters accord their support to the most credible proponent of an issue. Although much has been made of the stickiness of issue ownership (Petrocik 1996), more recent observations confirm that policy reputations are not static (Bélanger 2003). Through their campaign efforts, parties have reinforced or undermined linkages between political actors—themselves and others—and specific issue dimensions (Budge, Robertson, and Hearl 1987; Meguid 2002). Issue ownership, therefore, is subject to party manipulation.

I argue that a new conception of party strategies is needed, one that recognizes that parties compete by altering policy positions *and* the salience and ownership of issue dimensions. In the next section, I spell out the implications of this new conception of strategies for a theory of party competition between unequals.

## THE MODIFIED SPATIAL THEORY

### An Expanded Toolkit

In moving from a definition of strategies as purely programmatic tools to one with salience, ownership, and programmatic dimensions, our understanding of the range and effectiveness of party tactics increases. In contrast to spatial theories that emphasize party movement on a given issue dimension, this new theory suggests strategic behavior toward a niche party starts one step earlier—with the decision regarding *mainstream party entry*. Established parties must decide whether to recognize and respond to the issue introduced by the niche party. Party presence on a specific policy dimension, like the environment or immigration, is not a given.<sup>8</sup>

Parties finding an issue unimportant or too difficult to address can decide to ignore it. Rather than indicating a party's failure to react, this previously ignored "non-action" is a deliberate tactic that I call a *dismissive strategy*. By not taking a position on the niche party's issue, the mainstream party signals to voters that the issue lacks merit. If voters are persuaded that the niche party's issue dimension is insignificant, they will not vote for it. Thus, even though a dismissive strategy does not challenge the distinctiveness or ownership of the niche party's issue position, its salience-reducing effect will lead to niche party vote loss.

Conversely, parties can compete with the new party by adopting a position on its issue dimension. The salience of that issue increases as the mainstream party acknowledges the legitimacy of the issue and signals its prioritization of that policy dimension for electoral competition. Depending on the position that the mainstream party adopts upon entering the new issue space, this response is one of the already familiar accommodative (convergence) and adversarial (divergence) strategies.

Although both boost issue salience, the similarities between accommodative and adversarial tactics end there. An accommodative tactic undermines the distinctiveness of the new party's issue position, providing like-minded voters with a choice between parties. Consistent with standard spatial models, those voters closer to the accommodating mainstream party on the new issue will desert the niche party. But, according to my theory, even those voters who are (programmatically) indifferent between the two parties may be persuaded to leave the new party. By challenging the exclusivity of the niche party's policy stance, the accommodative mainstream party is trying to undermine the new party's issue ownership and become the rightful owner of that issue. The mainstream party is aided in this process by its greater legislative experience and governmental effectiveness. In addition, mainstream parties generally have more access to the voters than niche parties, allowing them to publicize their issue positions and establish name-brand recognition.<sup>9</sup> Given these advantages, the established party "copy" will be perceived as more attractive than the niche party "original."

In addition to strengthening the already powerful tool of convergence, the salience and ownership dimensions also empower the commonly ignored spatial strategy of policy divergence. When a party adopts an adversarial strategy, it declares its opposition to the niche party's policy stance. This strategic behavior calls attention to that challenger and its issue dimension, leaving voters primed to cast their ballots on the basis of this new issue. The adversarial strategy also reinforces the niche party's issue ownership by defining the mainstream party's issue position in juxtaposition

<sup>7</sup> For competition between unequals, this means that mainstream parties compete with the niche party using strategies restricted to the new issue dimension. This constraint allows us to avoid the problems of modeling competition between multiple players in multiple dimensions (Enelow and Hinich 1984).

<sup>8</sup> The work on party realignment does recognize that political actors might not take positions on all issue dimensions. However, even this body of research (e.g., Rohrschneider 1993) has not included the

decision to ignore new issue dimensions in its repertoires of party strategy.

<sup>9</sup> That exposure occurs through the media and the mainstream party's activists. The latter are typically more numerous and better integrated into society than those of the niche party.

**TABLE 1. Predicted Effects of Mainstream Party Strategies (in Isolation)**

Strategies	Mechanism			Niche Party Electoral Support
	Issue Salience	Issue Position	Issue Ownership	
Dismissive (DI)	Decreases	No movement	No effect	Decreases
Accommodative (AC)	Increases	Converges	Transfers to mainstream party	Decreases
Adversarial (AD)	Increases	Diverges	Reinforces niche party's ownership	Increases

**TABLE 2. Predicted Effects of Mainstream Party Strategic Combinations on Niche Party (NP) Electoral Support**

Mainstream Party A	Mainstream Party B		
	Dismissive (DI)	Accommodative (AC)	Adversarial (AD)
Dismissive (DI)	NP vote loss	NP vote loss	NP vote gain
Accommodative (AC)	NP vote loss	NP vote loss	If AC > AD, NP vote loss If AD > AC, NP vote gain
Adversarial (AD)	NP vote gain	If AC > AD, NP vote loss If AD > AC, NP vote gain	NP vote gain

to that of the new party. It strengthens the link in the public's mind between that issue stance and the niche party as its primary proponent. As a result, the adversarial strategy encourages niche party electoral support.

The predicted effects of this expanded set of party strategies on issue salience, ownership, and party programmatic position and, in turn, niche party vote are summarized in Table 1. Given that a niche party's support depends on a single issue, any tactic that undermines the perceived relevance of that issue, or the distinctiveness or credibility of the niche party's position on that dimension will result in vote loss. Assuming that voters find the niche party's policy stance attractive, mainstream parties can undermine niche party vote with dismissive or accommodative tactics and boost it with adversarial strategies.

### Changing the Nature of Party Competition: The Critical Role of Non-Proximal Parties

The expanded conception of strategies alters our understanding of the range and effectiveness of political tactics. But the implications of this revision extend far beyond the size of the party's toolkit. They call into question the very rules of party interaction propounded by spatial models. Recall that in the standard spatial conception of strategy, parties can only affect the electoral support of neighboring parties; in a unidimensional space, this means that movement by a center-left party away from a center-right party cannot impact the electoral support of a right flank party. If instead strategies can also alter issue salience and ownership, then parties can target opponents *anywhere* on that dimension. Ideological proximity is no longer a requirement.

Consider the effects and utility of the adversarial strategy. Given that political opponents are generally viewed as threats, it might seem counterintuitive to suggest, as I did, that a party would seek to heighten the

visibility and electoral strength of a competitor. Indeed, in a two-party system where politics is a zero-sum game, political parties are unlikely to employ adversarial tactics. When competition occurs between three or more players on a single dimension, however, such a vote-boosting strategy might be used against a competitor on the opposite flank of the political spectrum. The salience- and ownership-altering aspects of adversarial tactics allow mainstream parties who are not directly threatened by the niche party to use it as a weapon against their mainstream party opponents. This is the political embodiment of the adage "the enemy of my enemy is my friend"; the mainstream party helps the niche party—the enemy of its enemy in this case—gain votes from the other mainstream party. As this discussion intimates, failure to consider the tactics of the non-proximal party could lead to faulty predictions about niche party support.

### Hypotheses of the Modified Spatial Theory

Table 2 contains the predictions of my modified spatial theory of party competition for niche party success. These hypotheses are based on the behavior of multiple mainstream parties on one dimension—the niche party's new issue dimension. For ease of presentation, I assume that there are only three parties in the political system—mainstream party A, mainstream party B, and the niche party.<sup>10</sup> Because the effect of each tactic is theorized to be independent of the identity of the strategizing mainstream party, six distinct strategic combinations emerge: DIDI, DIAC, DIAD, ACAC, ACAD, and ADAD. The predictions recorded in Table 2 represent the combined effects of each of the individual tactics from Table 1 on niche party support.

The reconceptualization of party strategies has a profound impact on the expected outcomes of party

<sup>10</sup> This restriction does not represent an intrinsic limitation of the model

competition. Not only does each party have multiple means of undermining and bolstering a niche party's vote, but also the electoral fortune of that niche party is shaped by the behavior of multiple mainstream parties. The predictions in Table 2 suggest that one party's behavior alone is rarely determinative of niche party support; mainstream parties can use tactics to thwart the strategic efforts of their mainstream competitor. For example, I posit that mainstream party B's adversarial (AD) strategy will decrease the effectiveness of mainstream party A's vote-reducing dismissive (DI) and accommodative (AC) tactics. In the case of a DIAD combination, the salience, ownership, and programmatic effects of the adversarial strategy are expected to overpower the simple salience-reducing impact of the dismissive strategy. The result will be a more popular niche party with strengthened issue ownership.

The expected outcome of the ACAD strategy is less straightforward. Although the adversarial behavior of mainstream party B prevents its mainstream opponent from easily coopting the niche party's issue ownership and issue voters, B's ability to bolster the neophyte's vote depends on the relative intensity of the two strategies, where intensity is a function of the prioritization, frequency, and duration of party tactics. In this situation, best described as a battle of opposing forces, the mainstream party employing the greater number of tactics consistently for the longer period of time will prevail. If the accommodative strategy is more intense than the adversarial one, I expect that the niche party will lose issue ownership and issue-based voters to the accommodating party. If the adversarial tactic is stronger and more consistently employed, then the issue ownership of the niche party will be strengthened, and its electoral support will increase.

The effectiveness of these strategic combinations is not without constraints, however. Mainstream party tactics must be accompanied by changes in voters' perceptions of party positions, issue salience, and issue ownership. As in all theories of strategic interaction, policy inconsistency limits the success of a party's strategy; the promotion of contradictory policy stances either simultaneously or over time raises doubts among the voters about the credibility of the strategizing actor. My reconception of strategies as issue-ownership-altering devices also means that the utility of these tactics depends on their implementation shortly after the emergence of the niche party on the electoral scene. Once the voters identify the niche party as the sole proponent of the issue, the costs involved in undermining that perceived ownership render its likelihood slim. Hesitation, therefore, undermines the potency of these reconceptualized strategies.

## DATA

### Dependent Variable

To test the hypotheses of my modified spatial theory, I look at the electoral trajectories of niche parties that

emerged and contested national-level legislative elections in Western Europe from 1970 to 2000. The dependent variable is operationalized as the percentage of votes received by a given niche party in a national legislative election.<sup>11</sup>

In order to best examine the success of these parties across the entire set of Western European countries, my analysis focuses on the most common set of niche parties: the environmental and radical right parties. Following from my original description of niche parties, I categorize individual parties on the basis of their primary issue positions. Those single-issue actors prioritizing the environment are labeled green parties, and those emphasizing issues of law and order and immigration are deemed radical right parties. The resulting categorization is largely consistent with the classifications made by other party researchers (e.g., Golder 2003; Kitschelt 1994).

Given that mainstream party strategies are implemented only after new party challengers have developed, the cases in this analysis are limited to those instances of green and radical right party emergence.<sup>12</sup> Even with this restriction, the resulting set of niche parties provides a larger and more diverse set of cases than those examined in previous single-issue party analyses (e.g., Golder 2003). As summarized in Appendix Table A1, the dataset covers the electoral trajectories of 30 single-issue parties across 17 Western European countries: all green and radical right parties contesting multiple national legislative elections, regardless of their peak vote level.<sup>13</sup> Their electoral trajectories are examined from 1970 to 2000, a period that encompasses the life spans of the majority of these niche parties to date.

### Independent Variables

**Mainstream Party Strategies.** I argue that the competitiveness of niche parties is directly shaped by the behavior of their fellow political contestants. Although the political arena may contain up to 50 party competitors in any one national legislative election, this analysis focuses on the tactics of a subset of political actors:

<sup>11</sup> With the data organized as niche party panels, the separate inclusion of multiple green or multiple radical right parties from the same country would violate the assumed independence of the observations; it would introduce the possibility that the electoral failure of a green party simply reflects the success of a different green party in that country. Thus, for those countries in which two or more green parties contest a given election, the value of the dependent variable for that country-party-election observation is the sum of those parties' votes. The same adjustment is made for countries with multiple radical right parties.

<sup>12</sup> This is different from sociological models in which observed rates of unemployment can be used to impute latent green or radical right party support in the absence of party formation (Golder 2003; Jackman and Volpert 1996; Swank and Betz 2003). This article, therefore, assesses the impact of the explanatory variables on niche party vote conditional upon niche party entry.

<sup>13</sup> This requirement led to the elimination of the eighteenth country—Iceland—from the analysis. The Icelandic green party, *Vinstrihreyfingin—grænt framboð*, only contested one national-level election during the time period under examination.

the mainstream parties of the center-left and center-right.<sup>14</sup> Defined by both their location on the Left-Right political dimension and their electoral control of that Left or Right ideological bloc, mainstream parties are typically governmental actors. As discussed previously, their name recognition, media access, and status as governmental players provide them with strategic tools unavailable to smaller, less prominent political parties.

Mainstream parties from the 17 countries were initially chosen according to their position on the Left-Right axis. Drawing on the party classification structure proposed by Castles and Mair (1984, 83), mainstream parties of the center-left, or "Moderate Left," were defined as those parties with scores of 1.25 to 3.75 on a scale of 0 to 10. Mainstream parties of the center-right, Castles and Mair's "Moderate Right" parties, were those parties with positions of 6.25 to 8.75.<sup>15</sup> Where more than one party met the same criterion in any given country, the party with the highest electoral average from 1970 to 2000 was chosen. This system yielded one mainstream center-left and one mainstream center-right party in each country, with one exception: Ireland was recognized as having two center-right parties.<sup>16</sup> The resulting classifications are consistent with the rank ordering of parties reported in Laver and Hunt (1992). The mainstream parties included in the study are listed in the Appendix.

I drew on data from the Comparative Manifesto Project (CMP) to determine mainstream parties' responses to the niche parties. This dataset records a party's support for and prioritization of a set of issue positions.<sup>17</sup> Recall that, although niche parties introduce a new dimension to a political arena already defined by other issues, mainstream party strategies toward the new party are restricted to the one new dimension. Based on CMP measures of party policy related to the new issue axes, therefore, I coded the strategies of individual mainstream parties as dismissive, accommodative, or adversarial.<sup>18</sup> Support for law

and order (variable 605), a national(istic) way of life (601), and traditional morality (603) and opposition to multiculturalism (608) were deemed indicative of mainstream party accommodation of radical right parties.<sup>19</sup> Mainstream adversarial tactics were signaled by opposition to a national(istic) way of life (602) and traditional morality (604) and support for multiculturalism (607).<sup>20</sup> Environmental protection (501) and anti-growth economy (416) explicitly mention support for the environment; manifesto coverage of these topics was considered reflective of mainstream party accommodation of green parties. In the absence of any variable recording opposition to environmental protection, I used support for free enterprise (401) and agriculture and farmers (703) and opposition to internationalism (109) to capture adversarial strategies toward green parties. A party neither supporting nor opposing a niche party's issue, as indicated by the presence of little to no discussion of that topic in its election manifesto, was categorized as engaging in dismissive behavior. This coding procedure was conducted for each mainstream party for each national-level election between 1970 and 2000.<sup>21</sup> To ensure their validity, the resulting coding decisions were checked against mainstream party policy deliberations and pronouncements recorded in archival materials, contemporaneous news sources, and secondary analyses.<sup>22</sup>

From the classification of individual mainstream party tactics, I find occurrences of each of the six possible strategic combinations in the data. I model DIDI, DIAC, DIAD, ACAC, and ADAD as simple dummy variables. The effect of the sixth strategic combination, ACAD, depends on the relative intensity of the constituent strategies, with intensity measured by the percentage of each party's manifesto devoted to its issue position. I code the ACAD variable -1 when the intensity of the AC strategy is greater and +1 when the intensity of AD is greater.

As currently operationalized, the strategic variables capture mainstream party behavior toward niche parties independent of the tactics the mainstream parties employed in previous electoral periods. However, my modified spatial model posits that policy inconsistency and delay can undermine strategic effectiveness. A

<sup>14</sup> In results not presented here, I find that the addition to the model of variables capturing the strategic behavior of a third set of mainstream parties—the centrist parties—does not change the results. When the strategic responses of center-left and center-right mainstream parties are controlled for, centrist party tactics generally prove insignificant.

<sup>15</sup> With an average score of 5.4, Italy's commonly recognized center-right party, *Democrazia Cristiana*, was the exception. See Castles and Mair 1984, 80.

<sup>16</sup> The dominance of a noneconomic dimension in Irish politics means that *Fianna Fáil* and *Fine Gael* are largely indistinguishable on the Left-Right spectrum.

<sup>17</sup> Though there is disagreement in the literature as to whether precise spatial positions can be derived from CMP data, it is not necessary to join that debate here; information about the precise spatial position of a mainstream party on a particular issue is not necessary for my coding of party behavior.

<sup>18</sup> These measures of strategy capture the policy behavior of parties, not the effects of those tactics on voter perceptions of the salience and ownership of the niche party's issue. Because the predictions of the standard and modified spatial theories are not observationally equivalent, conclusions about the relative explanatory power of these strategic theories can be drawn without looking at the micro-level mechanism. In case studies of mainstream party-niche party inter-

action conducted elsewhere (Meguid 2002), I find direct evidence supporting my hypotheses about the issue salience- and ownership-altering mechanisms of strategies.

<sup>19</sup> For a strategy to be coded accommodative, a party's pronounced support of a neophyte's issue position could be accompanied by few references in opposition to that policy stance. A similar confirmatory procedure was employed when coding the adversarial tactics.

<sup>20</sup> The Comparative Manifesto Project does not include a negative corollary to Variable 605 measuring support for law and order. As noted by Laver and Garry (2000, 621), not all issues coded in the dataset are presented as positional issues—topics with positive and negative stances to them.

<sup>21</sup> Manifestos for a particular national-level election reflect the strategies adopted by mainstream parties sometime after the previous election but before the one being contested.

<sup>22</sup> The following resources were examined: British Labour and Conservative Party Archives; French Socialist Party Archives; Hainsworth 2000; Keesing's Worldwide 1999; Kitschelt 1994, 1995, O'Neill 1997; and Taggart 1996.



review of mainstream party strategies in my dataset reveals that policy hesitation, more than policy inconsistency, occurs during mainstream party–niche party interaction in Western Europe. Of the 114 observations, there are 18 cases of mainstream parties employing accommodative tactics (ACAC or DIAC) following two or more periods of dismissive strategies. Because there are only two instances of a mainstream party switching between AC and AD tactics in successive electoral periods in my data, I model only policy delay. I create time-sensitive dummy variables for DIAC and ACAC strategies, where the variables are coded 1 when the strategy was implemented after two or more periods of joint dismissive tactics. Because a mainstream party's ability to acquire issue ownership—the key mechanism behind accommodative tactics—is limited once voters deem the niche party the sole issue owner, hesitation is expected to counteract the vote-reducing power of these strategies.

**Institutional and Sociological Variables.** In addition to these strategic variables, I include those institutional and sociological factors identified by previous research as relevant to new party success. The permissiveness of the electoral and political environment is captured by two variables, a measure of district magnitude and a dummy variable indicating a centralized (as opposed to a federal) state structure.<sup>23</sup> Following the practices of Amorim Neto and Cox (1997) and Golder (2003), the first variable is operationalized as the logged magnitude of the median legislator's district.<sup>24</sup> The expectation is that, as district magnitude increases, niche party support will increase, with the marginal effect decreasing as the district magnitude becomes large. The second variable, state structure, is included to test the claim by Harmel and Robertson (1985) and Willey (1998) that the existence of subnational elected offices increases the electoral support of third parties at the national level.<sup>25</sup> As the variable is operationalized in this analysis, we expect a negative relationship; niche party vote levels should be lower in centralized than in federal systems.

To assess the significance of the sociological climate for niche party support, I use two measures of economic health: the current level of GDP per capita and the current rate of unemployment.<sup>26</sup> Unlike the effect of institutional variables, the predicted effect of these economic factors varies by niche party family. Green party vote is expected to be positively correlated with GDP per capita and negatively correlated with un-

employment (Taggart 1996). The relationships are the opposite for radical right party support (Golder 2003; Jackman and Volpert 1996). To allow for these party-specific effects, I model the economic variables as a series of party-specific terms.

Measures of postmaterialism and immigrant concentration—additional sociological measures of green (Inglehart 1998) and radical right party support (Golder 2003; Swank and Betz 2003)—were excluded from the model because of severe data restrictions and the lack of suitable proxies;<sup>27</sup> inclusion of these measures in the regression reduced the effective number of observations by half.<sup>28</sup> Although the significance of these variables cannot be tested against the full set of niche party observations, their inclusion in analyses with a reduced set of cases yielded nonsignificant coefficients and did not affect the substantive and statistical significance of the strategic variables.

## MODEL AND ANALYSIS

To estimate the effect of these institutional, sociological, and strategic factors on niche party electoral support, I employ pooled cross-sectional time-series analysis. Specifically, I ran an ordinary least-squares (OLS) regression with a lagged dependent variable, panel-corrected standard errors (Beck and Katz 1995, 1996), and country-fixed effects. The result of a joint F-test supports the inclusion of country dummy variables. Not only do these variables help to minimize country-level heteroskedasticity, which is not addressed by the niche party panel-level standard error correction of the model, but also they reflect country differences unaccounted for by the independent variables. These differences include, most importantly, variation in the distribution of voters' positions in the policy space—a variable for which no cross-country measure exists, yet which is critical to the predicted effect of mainstream party strategies on niche party support. As recommended by Beck and Katz (1995, 1996), the lagged dependent variable was added to eliminate autocorrelation in the underlying data.

## Findings

The regressions results are reported in Table 3, with the predicted signs of the independent variables listed

<sup>23</sup> Information on state structure was obtained from Harmel and Janda 1982, 72; and Elazar 1994.

<sup>24</sup> Data from Golder 2003.

<sup>25</sup> The logic of their claim is as follows. a decentralized system increases the number of representative positions and, thus, the likelihood that a new party will attain office. New parties who can draw on local-level governmental experience and grassroots support will gain higher vote shares when seeking national office.

<sup>26</sup> GDP per capita, reported at current prices and current purchasing power parity (PPP) in thousands of U.S. dollars, and unemployment, measured as a percentage of the total labor force, were taken from the *OECD Statistical Compendium*. CD-ROM 2000.

<sup>27</sup> The demographic variables typically associated with postmaterialist values—age and education—are not appropriate substitutes for the value orientation variable. Although age is negatively correlated with postmaterialism and green party support, it is also negatively correlated with materialist values and radical right support (Taggart 1996). Education has been found to have no relationship with green party vote when other factors are taken into account (Bürklin 1984).

<sup>28</sup> The Eurobarometer surveys only provide time-series data on postmaterialism for 11 of the 17 countries (*European Communities Studies*, 1970–1992), whereas the three waves of the *World Values Survey* only provide one observation per country per decade for a limited number of these countries. Similarly, data on the percentage of immigrants in a country are unavailable for nine elections across five countries in my dataset (Golder 2003).

**TABLE 3. Multivariate Analysis of Niche Party Vote Percentage**

Independent Variables	Predicted Sign	Niche Party Vote %
<b>Strategic</b>		
Mainstream Party		
DIDI	—	−1.37* (0.73)
ACAC	—	−1.52* (0.93)
DIAC	—	−0.92 (0.75)
DIAD	+	3.72* (1.84)
ADAD	+	6.54*** (1.56)
ACAD with relative intensity <sup>a</sup>	+	1.12* (0.59)
ACAC* hesitation	+	1.18 (1.06)
DIAC* hesitation	+	0.26 (1.55)
<b>Past performance</b>		
NP vote <sub>t-1</sub>		0.58*** (0.12)
<b>Institutional</b>		
Ln of median district magnitude	+	−0.51 (0.51)
State structure	—	−2.76* (1.34)
<b>Sociological</b>		
GDP/capita by niche party (in thousands)		
Green party	+	0.06 (0.05)
Radical right party	—	0.06 (0.06)
Unemployment by niche party		
Green party	—	−0.18** (0.07)
Radical right party	+	−0.14* (0.10)
Country dummies		Included
Adjusted R <sup>2</sup>		0.8656
N		114

Note \*\*\*  $p < .001$ ; \*\*  $p < .01$ ; \*  $p < .1$  (one-tailed tests based on panel-corrected standard errors). Standard error in parentheses. Analysis conducted using STATA 8.0

<sup>a</sup> The coefficient of the variable ACAD with Relative Intensity is reported in terms of the adversarial strategy being stronger than the accommodative one. Where  $AC > AD$ , the sign of the coefficient is the opposite

in column 2. The statistical significance of the coefficients is measured with one-tailed t-tests due to the directional nature of the institutional, sociological, and strategic hypotheses. For ease of presentation, the estimates of the 17 country dummies are not shown.<sup>29</sup>

The regression results confirm that the electoral trajectories of niche parties are not solely deter-

mined by—or, in some cases, even critically influenced by—institutional and sociological factors. Rather, mainstream party tactics exert statistically and substantively significant effects on niche party vote across elections. Of the factors used to test the competing institutional and sociological hypotheses, only the measures of state structure and unemployment in green party cases are significant and correctly signed predictors of niche party vote. Although insignificant findings could be encouraged by the lagged dependent variable model, which measures short-term determinants of niche party support levels, the statistical significance of the other institutional and sociological factors does not increase when the lagged dependent variable is dropped. The results are also robust to alternate specifications of the institutional and sociological variables.<sup>30</sup> Regardless of the configuration of the model and its variables, therefore, mainstream party action emerges as the central factor shaping niche party vote.

Beyond supporting the significance of strategic behavior, the analysis confirms that mainstream parties can use strategies either to weaken or to strengthen niche party electoral support. Consistent with the predictions of my modified strategic model, joint dismissive (DIDI) and joint accommodative (ACAC) tactics decrease, and dismissive–adversarial (DIAD) and joint adversarial (ADAD) tactical combinations increase, niche party support. The effect of dismissive–accommodative (DIAC) tactics, on the other hand, proves statistically insignificant. As expected, the impact of accommodative–adversarial (ACAD) tactics depends on the relative intensity of the constituent strategies. When adversarial tactics are dominant ( $ACAD = +1$ ), this strategic combination leads to an increase in niche party vote. When accommodative actions are stronger ( $ACAD = -1$ ), niche party support declines.

Although the effects of these strategies largely match the predictions of my modified spatial model, the regression results offer surprisingly little support for the claim that hesitation mitigates the vote-reducing power of accommodative tactics. A visual inspection of the data confirms that niche party support changes by a larger *positive* amount following the use of “delayed”—as opposed to timely—joint accommodative (ACAC) and dismissive–accommodative (DIAC) strategies, yet these vote-boosting effects do not appear to be significant when other factors are accounted for.<sup>31</sup> As this finding may be driven by the particular specification of the model, more attention to the potentially

<sup>30</sup> Replacement of the logged median district magnitude variable with alternate specifications—including the logged average district magnitude and a dummy variable capturing the plurality-proportional representation dichotomy—did not alter the results. Similarly, use of lagged economic variables did not change the significance of the sociological variables or any of the other variables in the model.

<sup>31</sup> Without controlling for other factors, the mean change in niche party vote following the timely implementation of ACAC is 0.61. It increases to 0.97 following the use of a “delayed” ACAC tactic. Similarly, mean change in niche party vote following the timely implementation of DIAC is 0.59. When the strategy is implemented after a delay, the mean change increases to 1.31.

<sup>29</sup> In all but one case—Spain—the coefficients of the country dummy variables were statistically significant at  $p < 0.1$  in a two-tailed test. Although these variables were included to account for unmeasurable country-level characteristics, the sign and magnitude of the specific country coefficients are not, in and of themselves, of interest here.

**TABLE 4. Predicted versus Observed Effects of Strategies on Niche Party Vote: Assessing the Standard Spatial Model's Predictions**

Predicted Effects on Vote According to Std. Spatial Model	Strategies	Observed Effects on Vote (Coefficients from Table 3)	90% Confidence Intervals
?	Dismissive	{ DIDI	−1.37 −2.57 to −0.17
Decrease	Accommodative	{ ACAC	−1.52 −3.05 to 0.00
		{ DIAC	−0.92 −2.15 to 0.31
		{ ACAD AC>	−1.12 −2.09 to −0.15
		{ ACAD AD>	+1.12 0.15 to 2.09
Increase	Adversarial	{ DIAD	+3.72 0.70 to 6.75
		{ ADAD	+6.54 3.97 to 9.11

Note: In the second column, the mainstream party strategic combinations are grouped by the tactic of the party proximal to the niche party on its new issue dimension

confounding effects of hesitation is necessary in future studies.

On the whole, then, the regression results provide support for my strategic model of niche party success. Do they, however, contradict the claims of the traditional spatial theories of party interaction? Can we conclude that strategies follow a micro-level mechanism whereby tactics alter issue salience and ownership, not just party programmatic position? To help answer these questions, I have summarized in Table 4 the expected impact of each strategy in a unidimensional space according to the standard spatial model along with the strategy's observed effect from the regression. Recall that the actions of a non-proximal party are considered irrelevant by the standard spatial theory. Thus, if we focus only on the behavior of the mainstream party closest to the niche party on this new issue, we can reduce our set of six different strategic combinations to three: those where there are no proximal parties (DIDI), those where the proximal party is accommodative (ACAC, DIAC, ACAD where  $AC > AD$ , and ACAD where  $AD > AC$ ) and those where the proximal party is adversarial (DIAD and ADAD).<sup>32</sup> These three strategic groupings are presented in column two of Table 4.

A comparison of the predicted and observed effects of these strategies offers some support for the standard spatial model. As anticipated by that theory in unidimensional competition, adversarial tactics employed by the proximal party—represented by DIAD and ADAD in our set of mainstream party responses—lead to neophyte vote gain. Accommodative strategies, in general, also have the expected effect—niche party vote loss. The standard spatial theory offers no clear predictions about the impact of dismissive tactics, or

not taking a position along the new policy dimension, on target party vote levels.

But the shortcomings of the standard spatial model begin to surface when we compare the effects of strategies within each of these three categories. No two of the combinations containing accommodative strategies have regression coefficients of the same magnitude. Greater inconsistencies emerge when we compare the coefficients of the strategic combinations in which the proximal party is adversarial. Consideration of the confidence intervals around these point estimates reduces the perceived differences between the strategies within each of the three categories, but several discrepancies remain. When accommodative tactics are paired (i.e., ACAC strategy), they reduce niche party support by 1.5 percentage points. Yet, when accommodation is joined with a more intense adversarial tactic (ACAD where  $AD > AC$ ), niche party vote *increases* by 1.1 percentage points.<sup>33</sup> The power of the “irrelevant” non-proximal party is also evident when we compare the effect of the accommodatively dominant ( $AC > AD$ ) and the adversarially dominant ( $AD > AC$ ) versions of accommodative–adversarial (ACAD) strategies. According to the standard spatial model, the effect of these strategies should be the same; yet there is a significant difference in niche party vote obtained after their implementation—a difference expected by my modified spatial model. These findings clearly demonstrate that the behavior of the *distant* party matters. Based on this comparison of the observationally distinct predictions of the two strategic theories, it seems that the logic of the modified spatial model captures competition between unequals better than that of the standard spatial model.

### FROM ONE ELECTION TO MANY: EXPLAINING A NICHE PARTY'S ELECTORAL TRAJECTORY

The regression parameter estimates confirm the central claim of my strategic model: mainstream party strategies affect the electoral strength of niche parties

<sup>32</sup> Proximity to the niche party in this unidimensional space is determined by the position adopted by a mainstream party upon entering the new issue dimension. A party acting accommodatively is proximal to the niche party on the new issue. The adversarial party is considered to be non-proximal unless no other mainstream party is accommodative; in that case, the adversarial party is considered proximal. Where both parties refuse to take a position on the new issue dimension (i.e., both act dismissively), there is no proximal party.

<sup>33</sup> The two-tailed 90% confidence intervals of these two strategic combinations do not overlap.

**TABLE 5. Electoral Trajectory of Radical Right Party**

Electoral Periods	Cumulative Electoral Support Level (%)	Change in Electoral Support (in Percentage Points)
Party Emerged: Base Level Vote ( <i>Exogenous to Model</i> )	3.00	
Period One: DIDI	4.96	+1.96
Period Two: DIAD	11.19	+6.23
Period Three: ACAD (where AD > AC)	12.23	+1.04
Period Four: ACAD (where AD > AC)	12.84	+0.61
Period Five: ACAD (where AC > AD)	10.95	-1.89

*Note:* Values calculated for a centralized state with a plurality electoral system and sociological variables held constant at their means. The French country dummy variable is coded 1

in a given election. But what can this model tell us about the shape of a niche party's electoral trajectory when mainstream parties employ different strategies over time? In Table 5, I present a typical set of mainstream party responses to a radical right party and its estimated effect on that niche party's vote over several elections. In this example, the electoral support levels are evaluated under plurality-based electoral rules in a centralized state with all economic variables held constant at their means. A vote of 3% was chosen to represent the niche party's opening electoral performance.

Although a mainstream party's initial strategy is contingent on a neophyte's degree of electoral threat, a survey of the data shows that most implement a cautious, low-cost dismissive tactic in the first electoral period.<sup>34</sup> Following the second electoral showing of the niche party, mainstream parties often take more active measures. Here a dismissive-adversarial tactic is modeled, a combination which more than doubles the vote level of the radical right party and transforms it from a minor irritant into a significant electoral threat. Tempering the effect of adversarial strategies with accommodative ones slows the rate of new party electoral gain. However, it is clear that a reduction in the absolute level of neophyte support occurs only when the intensity of the mainstream party's cooptative tactics surpasses that of the vote-bolstering adversarial ones.

Far from being a mere hypothetical, this scenario resembles the set of strategies pursued by the French mainstream Socialist (PS) and Gaullist (RPR) parties against the radical right Front National (FN) from 1978 to 1997. The Socialist party adopted an early, adversarial stance against the niche party. The internally divided Gaullist Party, on the other hand, was slow to respond actively to the threatening anti-immigrant party; the RPR pursued a cooptative strategy only as of 1986, *after* the electoral and reputational entrenchment of the FN. In contrast to the hypothetical presented above, the RPR's accommodative strategy remained weaker than the PS's adversarial tactics throughout this time period.

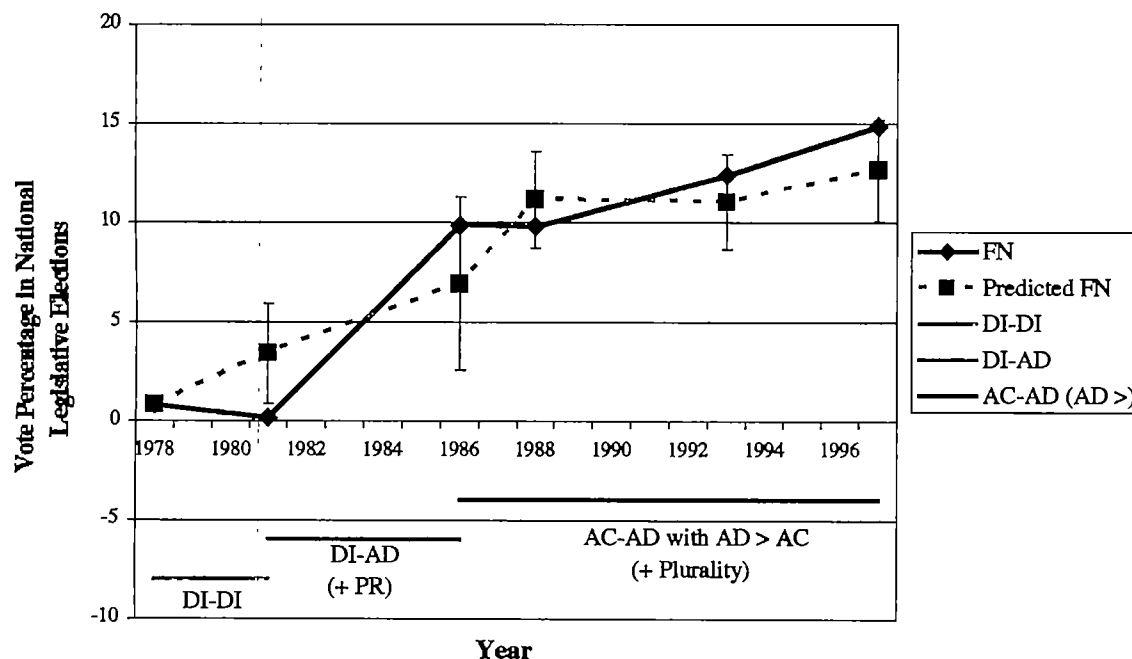
A comparison of the predicted effects of these mainstream party strategies with the niche party's actual electoral trajectory demonstrates the explanatory

power of my model. In Figure 1, I plot these trajectories, with the model's predictions of FN support from 1981 to 1997 based on the set of mainstream party strategies, institutional and sociological conditions, and lagged FN vote observed in France. As the Figure reveals, in four of the five predicted elections, the 95% confidence intervals around each point estimate encompass the actual FN vote share. Although we cannot fully ignore how GDP per capita and unemployment rates varied during this time period, the significant electoral gains made by the FN cannot be attributed to these sociological variables; in each of these elections, the joint effect of the sociological variables was to depress—not to boost—the vote share of the niche party. Thus, it was the strategic maneuvering of the French Socialists and Gaullists that served as the workhorse of the FN's electoral success.

In addition to confirming the power of mainstream party strategies, this comparison also calls attention to the role played by each established party in altering niche party success. In my model, increases in niche party support came largely at the hands of a traditionally ignored, non-proximal party. This prediction is consistent with the facts in the French case: it is readily accepted by French scholars and journalists that the FN's high vote percentages were the direct result of the PS's adversarial behavior (Faux, Legrand, and Perez 1994).<sup>35</sup> Being the "enemy of the PS's enemy" proved electorally fruitful for the FN. On the contrary, the proximal Gaullist party was relatively ineffective at containing the radical right party's support. The vote-diminishing influence of its dismissive and accommodative tactics was repeatedly overwhelmed by the adversarial behavior of its Socialist counterparts. Had we assumed that meaningful interaction only occurs between proximal actors, as claimed by standard spatial models of party competition, we would have predicted FN electoral failure rather than its apparent success.

<sup>34</sup> For a discussion of the factors affecting mainstream party strategic choice, see Meguid 2002.

<sup>35</sup> The Socialist Party also engaged in institutional forms of adversarial strategy towards the FN. To boost the niche party's support to the detriment of the RPR, the PS changed the electoral rules from a two-ballot plurality formula to PR in 1986. The RPR reinstated the plurality formula in 1988. These changes to the electoral system are reflected in the predicted values in Figure 1.

**FIGURE 1. Electoral Trajectory of the French Front National: Actual versus Predicted (With 95% Confidence Intervals)**

Note: Predictions calculated for a centralized state with plurality rules and GDP/capita, unemployment levels and lagged FN vote as observed in France.

## CONCLUSION

By focusing on electoral rules, state structure, and the economic health and value orientation of a society, theories of new party electoral strength have prioritized the structure of the competitive arena over the behavior of the actors within it. The evidence presented here suggests that party strategies should not be overlooked. Across Western Europe, the strategies of the electorally and governmentally dominant parties shape the electoral fortunes of niche parties. Moreover, when the actions of the mainstream parties on the niche party's new issue dimension are taken into account, the standard institutional and sociological factors fail to exhibit a consistently significant effect on green and radical right party vote levels.

The findings also challenge the sufficiency of the standard spatial conception of party strategy. Additional data on voter perceptions of the salience and ownership of the niche parties' issues are needed to examine explicitly the micro-level mechanism behind party tactics, but the regression results reveal that mainstream parties competing with niche actors are not merely altering their positions along established policy dimensions with fixed salience. Rather, the results are consistent with a modified spatial logic, whereby mainstream parties also manipulate the salience and ownership of the new party's issue. It follows that competition is not restricted to interaction between ideological neighbors, as the standard spatial theory

claims; non-proximal parties play a critical role in the success and failure of Western Europe's niche parties.

In affirming the general hypotheses of my spatial theory, this analysis implies that mainstream party strategies influence more than just niche party vote. Indeed, competition between party unequals has ramifications for the long-run competition between mainstream party *equals*. First, mainstream party responses to the new parties change the effective dimensions of political competition. By adopting either an accommodative or an adversarial strategy, the mainstream party is prioritizing the niche party's issue dimension and including it within the mainstream political debate. Thus, not only is the shape of the policy space endogenous to party competition, but also the "success" of the niche party's issue is distinct from niche party electoral success. Immigration and the environment have become mainstream campaign topics in most Western European countries, even though many of the niche parties that introduced them have disappeared. Strategies directed against short-term threats, therefore, may have a lasting impact on the content of the political debate.

Second, in an even more direct manner, these strategies affect the very survival of the mainstream parties. When adversarial strategies are employed against a non-proximal niche party, they turn it into a weapon against an established party opponent. Even though mainstream party electoral success typically

depends on the party's attractiveness on multiple policy dimensions, such single-issue adversarial tactics have been responsible for the loss of mainstream party legislative seats and even governmental turnover. Examples are not restricted to Western Europe, as demonstrated by the role of the Republicans' adversarial tactics toward Green Party candidate Ralph Nader in the defeat of Democrat Al Gore in the 2000 U.S. presidential election. At the extreme, adversarial strategies could result in party system

realignment through the elimination of the mainstream party opponent and its replacement with the niche party. With consequences for both the number of parties and the issues dominating political debate, mainstream party tactics against niche parties are not just means to counteract a set of single-issue political actors; these everyday strategies have effectively become tools in the larger political processes of party system change.

## APPENDIX

**TABLE A1. Western European Mainstream and Niche Parties Included in the Analysis**

Country	Center-Left Mainstream Party	Center-Right Mainstream Party	Environmental Party	Radical Right Party
Austria	SPO	ÖVP	Die Grüne Alternative, VGO	FPÖ
Belgium	PS/SP	PRL/PVV	Ecolo/Agalev	Vlaams Blok, Front National
Denmark	SD	KF	De Grønne	Fremskridtspartiet
Finland	SSDP	KOK	Vihreat/Vihrea Liitto	—
France	PS	RPR	Les Verts, Génération Ecologie	Front National
Germany	SPD	CDU	Die Grünen	Die Republikaner, Deutsche Volksunion
Greece	PASOK	Nea Dimokratia	OIKIPA	—
Ireland	Labour	Fianna Fáil, Fine Gael	Comhaontas Glas	—
Italy	PCI	DC	Liste Verdi	Movimento Sociale Italiano
Luxembourg	LSAP	CSV	DI Grëng Alternative, Grëng Lëscht Ekologesch Initiative	Lëtzebuerg fir de Letzebuerger National Bewegung
Netherlands	PvdA	VVD	Groen Links, De Groenen	Centrumdemocraten, Centrumpartij / Centrumpartij '86
Norway	A	H	Miljøpartiet de Grønne	Fremskrittspartiet
Portugal	PSP	PSD	Os Verdes	Partido da Democracia Crista
Spain	PSOE	AP/PP	Los Verdes	—
Sweden	SAP	M	Miljöpartiet de Gröna	Ny Demokrati
Switzerland	SPS/PSS	CVP/PDC	Grüne Partei der Schweiz/ Parti écologiste suisse, Grünes Bündnis der Schweiz/ Alliance socialiste verte	Schweizerische Demokraten/Démocrates suisse, Schweizer Auto Partei/Parti automobiliste suisse
United Kingdom	Labour	Conservative	Green Party	National Front, British National Party

Sources: Caramani (2000); Castles and Mair (1984); Laver and Hunt (1992); Mackie and Rose (1991, 1997).



## REFERENCES

- Amorim Neto, Octavio, and Gary Cox. 1997. "Electoral Institutions, Cleavage Structures and the Number of Parties." *American Journal of Political Science* 41 (January): 149–74.
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (And Not to Do) with Time-Series-Cross-Section Data." *American Political Science Review* 89 (September): 634–47.
- Beck, Nathaniel, and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6: 1–36.
- Bélanger, Eric. 2003. "Issue Ownership by Canadian Political Parties since 1953." *Canadian Journal of Political Science* 36 (July/August): 539–58.
- Budge, Ian, and Denis Farlie. 1983. "Party Competition—Selective Emphasis or Direct Confrontation?" In *Western European Party Systems. Continuity and Change*, ed. Hans Daalder and Peter Mair. London: Sage, 267–306.
- Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, Eric Tannenbaum, et al. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford: Oxford University Press.
- Budge, Ian, David Robertson, and Derek Hearl. 1987. *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*. Cambridge: Cambridge University Press.
- Bürklin, Wilhelm. 1987. "Governing Left Parties Frustrating the Radical Non-established Left: The Rise and Inevitable Decline of the Greens." *European Sociological Review* 3 (September): 109–26.
- Caramani, Daniele. 2000. *Elections in Western Europe Since 1815*. New York: Macmillan.
- Castles, Francis G., and Peter Mair. 1984. "Left-Right Political Scales. Some 'Expert' Judgments." *European Journal of Political Research* 12 (March): 73–88.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper Collins.
- Duverger, Maurice. 1963. *Political Parties: Their Organization and Activity in the Modern State*. London: Methuen.
- Elazar, Daniel. 1994. *Federal Systems of the World*. Vol. 2. Harlow, UK: Longman Current Affairs.
- Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge: Cambridge University Press.
- Faux, Emmanuel, Thomas Legrand, and Gilles Perez. 1994. *La Main Droite de Dieu*. Paris: Editions du Seuil.
- Golder, Matt. 2003. "Explaining Variation in the Success of Extreme Right Parties in Western Europe." *Comparative Political Studies* 36 (May): 432–66.
- Hainsworth, Paul, ed. 2000. *The Politics of the Extreme Right*. New York: Pinter.
- Harmel, Robert, and Kenneth Janda. 1982. *Parties and Their Environments*. New York: Longman.
- Harmel, Robert, and John Robertson. 1985. "Formation and Success of New Parties: A Cross-National Analysis." *International Political Science Review* 6 (October): 501–23.
- Harmel, Robert, and Lars Svendsen. 1997. "The Influences of New Parties on Old Parties' Platforms." *Party Politics* 3 (3): 315–40.
- Inglehart, Ronald. 1998. *Modernization and Postmodernization*. Princeton: Princeton University Press.
- Jackman, Robert W., and Karin Volpert. 1996. "Conditions Favouring Parties of the Extreme Right in Western Europe." *British Journal of Political Science* 26 (October): 501–21.
- Keesing's Worldwide. 1999. *Keesing's Record of World Events CD-ROM*. Bethesda, MD: Keesing's Worldwide.
- Kitschelt, Herbert. 1994. *The Transformation of European Social Democracy*. Cambridge: Cambridge University Press.
- Kitschelt, Herbert. 1995. *The Radical Right in Western Europe*. Ann Arbor: University of Michigan Press.
- Laver, Michael, and John Garry. 2000. "Estimating Policy Positions from Political Texts." *American Journal of Political Science* 44 (July): 619–34.
- Laver, Michael, and W. Ben Hunt. 1992. *Policy and Party Competition*. New York: Routledge.
- Mackie, Thomas, and Richard Rose. 1991, 1997. *The International Almanac of Electoral History*. London: Macmillan.
- Mair, Peter. 1999. "New Political Parties in Established Party Systems: How Successful Are They?" In *Elites, Parties and Democracy*, eds. E. Beukel, K. Klausen, and P. Mouritzen. Odense: Odense University Press, 207–24.
- Meguid, Bonnie M. 2002. "Competition Between Unequals: The Role of Mainstream Party Strategy in Rising Party Success in Western Europe." Ph.D. diss. Harvard University.
- Muller-Rommel, Ferdinand. 1996. "The New Challengers: Explaining the Electoral Success of Green and Right-Wing Populist Parties in Western Europe." In *Les Petits Partis*, eds. Annie Laurent and Bruno Villalba. Paris: L'Harmattan, 119–41.
- Organization of Economic Cooperation and Development. 2000. *OECD Statistical Compendium CD-ROM*. Paris: OECD Electronic Editions.
- O'Neill, Michael. 1997. *Green Parties and Political Change in Contemporary Europe*. Brookfield, VT: Ashgate.
- Petrocik, John. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40 (August): 825–50.
- Rabinowitz, George, and Stuart Elaine Macdonald. 1989. "A Directional Theory of Issue Voting." *American Political Science Review* 83 (March): 93–121.
- Rohrschneider, Robert. 1993. "New Party Versus Old Left Realignments: Environmental Attitudes, Party Politics and Partisan Affiliations in Four Western European Countries." *Journal of Politics* 55 (August): 682–701.
- Shepsle, Kenneth A. 1991. *Models of Multiparty Electoral Competition*. New York: Harwood Academic Publishing.
- Swank, Duane, and Hans-Georg Betz. 1995. "Right-Wing Populism in Western Europe." Paper presented at the annual meeting of the American Political Science Association, Chicago.
- Swank, Duane, and Hans-Georg Betz. 1996. "Internationalization and Right-Wing Populism in Western Europe." Paper presented at the Conference on Globalization and Labor Markets, University of California, Los Angeles.
- Swank, Duane, and Hans-Georg Betz. 2003. "Globalization, the Welfare State and Right-Wing Populism in Western Europe." *Socio-Economic Review* 1 (May): 215–45.
- Taggart, Paul. 1996. *The New Populism and the New Politics*. New York: St Martin's Press.
- Wiley, Joseph. 1998. "Institutional Arrangements and the Success of New Parties in Old Democracies." *Political Studies* 46 (3): 651–68.
- Woldendorp, Jaap, Hans Keman, and Ian Budge. 1998. "Party Government in 20 Democracies: An Update (1990–1995)." *European Journal of Political Research* 33 (1): 125–64.

# Uncovering Evidence of Conditional Party Government: Reassessing Majority Party Influence in Congress and State Legislatures

WILLIAM T. BIANCO *Penn State University*

ITAI SENED *Washington University in St. Louis*

*This paper aims at enriching the debate over the measurement of majority party influence in contemporary American legislatures. Our use of a new analytic technique, a grid-search program for characterizing the uncovered set, enables us to begin with a better model of legislative proceedings that abandons the simple one-dimensional spatial models in favor of the more realistic two-dimensional version. Our conclusions are based on the analysis of real-world data rather than on arguments about the relative merits of different theoretic assumptions. Our analysis confirms that when legislators' preferences are polarized, outcomes will generally be closer to the majority party's wishes, even if the majority-party leadership does nothing to influence the legislative process. This conclusion notwithstanding, our analysis also shows that at the margin of the majority party's natural advantage, agenda setting by the majority party remains a viable and efficacious strategy.*

In the debate over party organizations in the American Congress, one side (e.g., Aldrich and Rohde 1998, 2001) argues that the majority party can influence the outcome of legislative proceedings through agenda control or the ability to determine which proposals are considered. The other side (e.g., Krehbiel 1999, 2000) argues that agenda control conveys no power and that the majority party's apparent influence stems from the fact that it has more elected members (thus more votes) than the minority party. In this paper we use new analytic techniques and real-world data to answer two questions that are at the heart of this debate: First, to what extent can majority-party leaders use power over the agenda to influence the results of legislative action? Second, under what conditions is this influence observable?

The dispute over majority party influence embodies a fundamental question about the factors driving legislative outcomes in both the modern Congress and the legislatures in general. Simply put, do parties matter? That is, if we are trying to explain why a particular proposal was enacted, defeated, or never even brought up for debate, must we consider agenda-setting efforts of majority party leaders as a potential explanatory variable? Alternatively, are legislative outcomes fully explained by what individual legislators are willing to vote for, with party leaders having no influence beyond the votes they cast as members of the chamber?

Resolution of this debate is important because the two theories embody very different predictions about the relationship between legislators' preferences and legislative rules on the one hand and policy outcomes on the other. If agenda control conveys an advantage to the majority party, then changes in which party holds majority status will generally alter outcomes (which policy is enacted), even if the overall distribution of legislators' preferences in the cham-

ber stays the same. Under this scenario, outcomes will also be sensitive to changes in the preferences held by majority-party legislators, changes in leaders' agenda power, and changes in the polarization of preferences between the majority and minority parties. If, on the other hand, control over the agenda is irrelevant—if parties “don't matter”—then changes in majority status, agenda power, or polarization will have no effect on legislative outcomes. Rather, outcomes will be sensitive to changes in the preferences of both majority and minority legislators. Thus, any attempt to explain legislative outcomes in the contemporary Congress requires a resolution of the debate over the role that party organizations and party leaders play in shaping these outcomes

Our contribution begins with a new technique for estimating the uncovered set, a concept that describes a fundamental constraint on legislative action: given the preferences of decision makers, reflecting personal taste and pressures ranging from constituent demands to progressive ambition, which outcomes can emerge from majority-rule decision making? In this paper, the uncovered set provides a baseline for assessing the potential for agenda setting, enabling us to move the debate over majority-party influence from a comparison of purely abstract models to a discussion framed in terms of actual preferences and feasible outcomes in real-world legislatures. We estimate uncovered sets for a number of U.S. House sessions and U.S. state legislatures, finding that the set of enactable outcomes in all these legislatures is relatively large and closer to majority-party legislators compared to those from the minority party. The degree to which enactable outcomes favor the majority party increases as a legislature becomes more polarized—as the difference in legislators' preferences across the majority and minority caucuses increases.

These results have two important implications. First, they confirm that observability concerns plague the measurement of party influence. Even if majority-party leaders make no attempt to shape the outcome of legislative proceedings, a comparison of outcomes with majority- and minority-party preferences will suggest

William T. Bianco, Department of Political Science, Pennsylvania State University, Ford Laboratory, University Park, PA 16802.

Itai Sened, Department of Political Science, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130

that majority-party leaders have successfully manipulated the legislative process. Second, even after accounting for this effect, our results show that a critical necessary condition for majority-party influence is met in many real-world legislatures: majority-party leaders can make themselves and their caucus better off by selecting one of many enactable outcomes and implementing agendas that yield this outcome from floor proceedings.

## THE UNCOVERED SET, PARTIES AND LEGISLATIVE OUTCOMES

The essence of formal modeling is to develop abstract representations of real-world situations and to use these models to predict real-world behavior and outcomes. Such models facilitate critical tests between competing theories. In the debate over party influence in the U.S. Congress, the technique of choice is spatial modeling, where preferences and outcomes are specified in terms of points in space.

How do preferences translate into outcomes in spatial voting games? In a one-dimension spatial model, the Median Voter Theorem (MVT) states that the expected outcome of majority-rule voting, given an open agenda and single peaked preferences, is the ideal point of the median voter (the core or Condorcet winner). Thus, if we know the median voter's ideal point, we can predict that it would be the outcome of any majority-rule voting in these games. Put another way, the median voter's ideal point is the only enactable outcome in the game. An outcome is enactable if there exists some admissible agenda such that (1) the outcome could be the ultimate result when sophisticated legislators vote through the agenda using majority rule and (2) the agenda itself could receive majority support when brought to the floor.

It is well known that the MVT does not generalize to cases where multiple dimensions are needed to describe preferences and outcomes, implying that outcomes in these games are sensitive to agendas, voting rules, and other constraints (Shepsle 1979, 1986). The so-called Chaos Theorems state that majority-rule decision making, unchecked by institutional constraints, can go "from anywhere to anywhere," rendering the ultimate outcome indeterminate. However, further work showed that if voters or legislators in these settings consider the ultimate consequences of their actions, rather than choose myopically between alternatives presented at each decision point, majority-rule voting will yield an outcome in a relatively small area, the uncovered set (McKelvey 1986, Miller 1980).

Formally, let  $N$  be the set of  $n$  voters or legislators. Assume that  $n$  is odd and for any agent,  $i \in N$ , preferences are Euclidian and defined by an *ideal point*  $\rho_i$ . Let  $x, y$  be elements of the set  $X$  of all possible outcomes. A point  $x$  *beats* another point  $y$  by majority rule if it is closer than  $y$  to more than half of the ideal points. A point  $x$  is *covered* by  $y$  if  $y$  beats  $x$  and any point that beats  $y$  beats  $x$ . The uncovered set includes all points not covered by other points. In essence, the uncovered

set generalizes the MVT to multidimensional spatial models. In a one-dimensional spatial model, the uncovered set is a single point—the median voter's ideal point. In a multidimensional game, the uncovered set is a relatively compact region within the space, unless ideal points satisfy stringent conditions in which a single core point exists and the uncovered is this core point.

The significance of the uncovered set lies in the potential to specify the set of possible majority-rule outcomes in these games and in the real-world situations these games intend to capture. If  $y$  covers  $x$ ,  $y$  dominates  $x$  as an outcome of a majority-rule voting game (McKelvey, 1986–8): if  $y$  covers  $x$ , any outcome that ties  $y$  defeats or ties  $x$  and any outcome that defeats  $y$  also defeats  $x$ . Therefore, strategic legislators should eliminate covered points from the voting agenda. Instead of promoting outcomes that are bound to be defeated later in the game, sophisticated legislators should promote points in the uncovered set that may survive the voting process (Cox 1987, 419).<sup>1</sup> Moreover, regardless of what "status quo point" a voting process may begin at, supporters of outcomes in the uncovered set can secure these outcomes using relatively simple (two-step) agendas and, moreover, defend them against opponents who propose outcomes outside the uncovered set (Shepsle and Weingast 1984). Thus, if we know which outcomes are in the uncovered set, we know what is possible in a legislative setting—which outcomes might be the ultimate result of legislative action.

Although much effort has focused on characterizing the size, shape, and location of the uncovered set in spatial games, a general result has eluded scholars up to now.<sup>2</sup> This paper utilizes a new grid-search computational method (Bianco, Jeliaskov, and Sened 2004a) for estimating the uncovered set for Euclidean preferences on a two-dimensional space. The approach follows McKelvey (1986, 27), treating the policy space as a set of discrete potential outcomes. It starts with two-dimensional preference data and compares points across the grid to determine the uncovered set's precise location, shape, and size.

The grid-search technique also allows us to determine whether the uncovered set's theoretic attractiveness is matched by an ability to predict real-world outcomes. Predictive power is crucial to our analysis: if the uncovered set does not capture actual outcomes, its size or location provides no insight into party influence or its observability. Bianco et al. (2004b) reanalyze data from canonical majority-rule experiments, showing that the uncovered set is a very good predictor of experimental outcomes—depending on the experiment,

<sup>1</sup> For similar arguments, McKelvey 1986; Miller 1980; Ordeshook and Schwartz 1987, and Shepsle and Weingast 1984, 1994.

<sup>2</sup> Previous analysis has identified four properties of the uncovered set: the uncovered set is never empty (McKelvey, 1986), if the core is nonempty, it coincides with the uncovered set (McKelvey 1986, Miller 1980); the uncovered set is a subset of the Pareto set (Miller 1980; Shepsle and Weingast 1984), and if  $r$  is the radius of the smallest ball  $Y$  that intersects all median hyperplanes, the uncovered set is contained within a ball of radius  $4r$  centered on  $Y$  (McKelvey)

over 90% (often 100%) of outcomes are in the uncovered set. Experiments using 5-player and 35-player groups (Bianco et al. 2004c) provide additional evidence of the uncovered set's predictive power. Bianco, Jeliaskov, and Sened (2004a) also show that, more often than not, the theoretically derived uncovered set is consistent with actual outcomes in the contemporary U.S. Congress.

### The Uncovered Set and the Observability of Party Influence

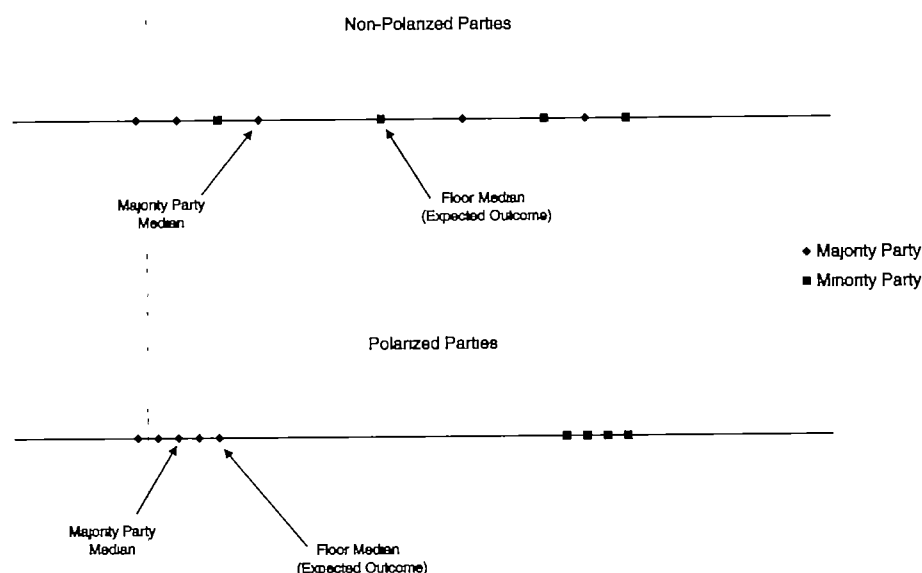
The theory of conditional party government (Aldrich 1995; Aldrich and Rohde, 1998, 2001) states that majority-party influence stems from its leaders' control of the agenda, debate, and institutions. When parties are polarized (similar policy goals within each caucus, disagreement across caucuses), "the majority party acts as a structuring coalition, stacking the deck in its own favor—both on the floor and in committee—to create a kind of "legislative cartel" that dominates the legislative agenda (Cox and McCubbins, 1993, 270; see also 2003)." As a result, "...the greater the degree of satisfaction of the condition in conditional party government, the farther policy outcomes should be skewed from the center of the whole Congress toward the center of opinion in the majority party" (Aldrich and Rohde 10–11).

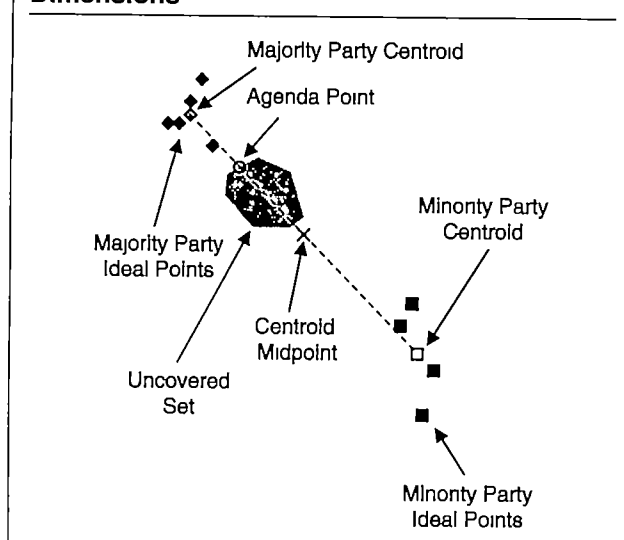
Some evidence supports conditional party government. Aldrich and Rohde (1998) find polarization in legislators' ideal points in the contemporary House. Aldrich and Battista (2000) explain this polarization in terms of electoral forces, finding a positive relationship between the effective number of parties in a state and various measures of legislative polarization. However, the key prediction of conditional party government—that outcomes will favor the majority party—has never been tested.

Krehbiel (1999, 2000) argues that outcomes favoring the majority party are a natural consequence of polarization, thus imply nothing about majority-party power: "Parties are said to be strong exactly when, viewed through a simple spatial model, they are superfluous (Krehbiel 1999, 35)." That is, when legislative parties are polarized, outcomes will lie closer to majority party ideal points because the party contains a majority of legislators, not because of anything party leaders do. This claim is derived from a unidimensional spatial model with two parties, where all legislators are party members. Thus, the median floor legislator is necessarily a majority party member. The MVT predicts that the outcome of this voting game will be the median voter's ideal point. Therefore, as the party medians diverge toward opposite ends of the dimension, the median voter—and expected outcomes—moves away from the center of the distribution and toward the majority party. However, this effect is the result of polarization coupled with majority rule; it does not require any actions by party leaders.

Figure 1 contains a nine-voter example of Krehbiel's argument. The figure contains two 1-dimensional spatial games where ideal points for the majority party are denoted by diamonds and for the minority party by squares. The top example depicts a situation where ideal points are not polarized. As the figure indicates, the expected outcome is the median voter's ideal point, located near the center of the plot. In the bottom example, ideal points are polarized with majority-party legislators on the left-hand side of the plot. The expected outcome is the median legislator's ideal point. However, with the change in the distribution of ideal points, the median legislator's ideal point, and, thus, the expected outcome, has shifted leftward. This shift, however, is driven by legislators' preferences, without any additional effect resulting from actions taken by party leaders.

FIGURE 1. Measuring Party Influence: The Problem of Observability In One Dimension



**FIGURE 2. Party Influence in Two Dimensions**

The concern about the observability of party influence is not a technical point. Conditional party government and Krehbiel's (1999) counterhypothesis are based on very different conceptions of the limits on majority-party action. The implicit assumption under conditional party government is that there is a wide range of enactable outcomes. The task for party leaders is to decide which of these outcomes are most acceptable to their caucus and to select procedures that generate this outcome. In contrast, in Krehbiel's model, there is only one enactable outcome—the median voter's ideal point. Thus, the majority party is fundamentally constrained by floor preferences. If majority-party leaders propose a nonmedian outcome or procedures that would lead to a nonmedian outcome, their proposal or agenda will not gain majority support.<sup>3</sup>

This debate provides an entry point for our analysis. Krehbiel's (1999) findings are based on a one-dimensional spatial model. The question is, does it generalize to more realistic two-dimensional situations? Our analysis begins with the premise that information on the size and location of the uncovered set in real-world legislatures offer a resolution to concerns about the observability of party influence—and a new scenario for the observable exercise of party influence. In particular, although the uncovered set is a single point in a one-dimensional game, it may well be larger in a more realistic multidimensional game. If so, the size of the uncovered set may create opportunities for agenda setting in multiple dimensions, even if it is offset toward the majority party as in a one-dimensional game. This scenario is shown in Figure 2.

Figure 2 gives a hypothetical configuration of ideal points and uncovered set in two dimensions for a

nine-voter, polarized party legislature. The figure labels the majority and minority centroids (the average ideal point for each party), draws a dashed line between centroids, and labels the midpoint of this line. The dark region is a hypothetical uncovered set, one that is typical of the real-world data presented later. Note that the uncovered set is relatively large and offset toward the majority party. In fact, it is completely on the majority party side of the line linking the party centroids. This location implies that actual outcomes, located inside the uncovered set, will always favor majority party legislators to some extent regardless of what their leaders do—just as in Krehbiel's (1999) one-dimensional model. However, the larger size of the uncovered set creates a new opportunity for the exercise of majority-party power. Rather than allowing outcomes to be determined by unconstrained floor action, majority-party leaders can pick a point inside the uncovered set, presumably one that is close to ideal points in their caucus, and use procedural strategies that yield this outcome on the floor. One such outcome is labeled as "agenda point" in Figure 2—note that all majority-party legislators prefer this outcome to everything else in the uncovered set. Suppose that party leaders structured the legislative process to yield this outcome. If an observer considered these outcomes and knew where the uncovered set was located, he or she would conclude that the distribution of preferences conveyed a natural advantage to majority-party legislators—but that majority-party power was being exercised at the margin of this advantage.

In sum, in a one-dimensional spatial model of legislative action, it is hard to see how the majority party could influence outcomes at all. However, this result may not hold in more realistic spatial models or in the real-world legislatures they depict. Our aim here is to derive uncovered sets for a variety of real-world legislatures using two-dimensional preference data and to assess their size and location relative to the majority and minority parties, allowing a direct resolution of the party influence debate.

More specifically, our focus is on two measurements. The first is a comparison of the average distance between outcomes in the uncovered set and the ideal points of legislators in the majority and minority parties. For some legislature, let  $M$  be the set of legislators in the majority party ( $m$  in number),  $N$  be the set of minority party legislators ( $n$  in number) and  $U$  the set of outcomes in the uncovered set ( $u$  in number). Let  $D_{ij}$  be the distance between legislator  $i$ 's ideal point,  $\rho_i$ , and an outcome  $j$  in the uncovered set. Let  $D_m$  ( $D_n$ ) denote the average distance between the ideal points of legislators in the majority (minority) party and outcomes in the uncovered set:

$$D_m = \sum_{i \in M, j \in U} D_{ij} / (m \cdot u),$$

$$D_n = \sum_{i \in N, j \in U} D_{ij} / (n \cdot u).$$

<sup>3</sup> The only exception is whether majority party leaders have a large supply of side payments (or punishments), enough to force a majority to support a nonmedian outcome, either in an up-or-down vote or in votes to establish institutional constraints that lead to the enactment of such an outcome.

We express the difference in these measures as a percentage:  $100 \cdot (D_n - D_m)/D_n$ . This measure varies from  $-100$  to  $100$ . If the uncovered set is equidistant from majority- and minority-party legislators, it will equal zero. If the uncovered set is closer to the majority party on average, as per Krehbiel (1999), the measure will be positive. For example, if the average distance between majority party legislators and uncovered set outcomes is half as large as the distance between these outcomes and minority party legislators, the measure will equal 50. Negative values imply that the uncovered set is closer to the minority party than to the majority party. Difference-of-means tests will be used to assess the statistical significance of the difference in average distance between the majority and minority parties.

Our second measurement focuses on the potential gains from agenda setting within the uncovered set. That is, to what extent can efforts to produce one particular uncovered set outcome improve on the situation where leaders are inactive and all uncovered set outcomes are equally possible? In general, it is not obvious which outcome will be the focus of party leaders' agenda-setting efforts. In real-world legislatures, this calculation is a complex process involving the preferences of caucus members; factional splits in the caucus; and the availability of side payments, threats, and strategic behavior on behalf of the relevant legislators. Our analysis approximates this calculation as follows. Rather than considering legislators' utilities or payoffs, we focus on distance: to what extent can agenda setting within the uncovered set bring outcomes closer to the ideal points of majority party legislators compared to the expected distance given no leadership action?<sup>4</sup>

We assess the potential for majority party agenda setting as follows. For a given legislature, let  $\rho_m = \sum_{i \in M} \rho_i / m$  denote the center of gravity for the majority party ideal points, where  $\rho_i = (x_i, y_i)$  is a two-dimensional vector denoting legislator's  $i$ 's ideal point. We interpret it as an approximation of the outcome that majority-party leaders and caucus members would like to enact if possible.<sup>5</sup> Let  $D_{mj}$  be the distance between  $\rho_m$  and an outcome  $j$  in the uncovered set. We calculate the average distance between  $\rho_m$  and all of the outcomes in the uncovered set and denote it by  $C$  so that:  $C = \sum_{j \in U} D_{mj} / (u)$ . This distance gives a baseline for how the majority caucus evaluates a situation where their leaders are inactive and all uncovered set outcomes are equally likely—on average, how far away are these outcomes from what caucus members would like to enact.

The next step is to determine how much party leaders can improve on this baseline. Let  $x \in U$  be the uncov-

ered set outcome that is closest to  $\rho_m$  and denote the distance between  $x$  and  $\rho_m$  by  $C_x$ . Outcome  $x$  is the best that party leaders can do in terms of using agenda setting to satisfy their caucus—anything closer to  $\rho_m$  is outside the uncovered set and, therefore, unenactable.

Our analysis of the potential for agenda setting in a given legislature focuses on the difference between  $C$  and  $C_x$ . We scale both distances by expressing them as a percentage of the range of legislator ideal points on the  $x$ -axis to normalize the results across different legislatures and data sources. In addition, scaling provides insight into the substantive significance of the two distances and the differences between them.<sup>6</sup> The theory of conditional party government would predict that a conditional party government is more likely insofar as  $C$  is relatively large, indicating that the set of possible outcomes (the uncovered set) is relatively large and contains many outcomes that are far away from what members of the majority caucus would like to enact. Moreover, the theory would predict that conditional party government is more likely to exist insofar as  $C_x$  is smaller than  $C$ , implying that the potential gains from agenda setting are substantial.

## UNCOVERED SETS AND CONDITIONAL PARTY GOVERNMENT

This section assesses the potential for conditional party government in contemporary legislatures by analyzing the size and location of uncovered sets for various sessions of the U.S. House of Representatives and state legislatures. Regarding our data, there is an ongoing and unresolved debate over the appropriate technique for recovering legislators' ideal points from observed behavior (votes).<sup>7</sup> As our goal is to analyze majority party influence rather than to adjudicate among the various techniques—and because there is no consensus about which is best—we use several datasets, with the aim of showing that our results are robust to heterogeneous and diverse data sources.<sup>8</sup>

<sup>6</sup> Whereas our focus here is on substantive significance, the agenda-setting results we present later are all statistically significant (difference of proportion tests) at the usual significance levels.

<sup>7</sup> A partial list of relevant papers includes Cox and Poole 2002; McCarty, Poole, and Rosenthal 2001; Londregan 1999; Poole and Rosenthal 2001; Groseclose and Snyder 2001, and Clinton, Jackman, and Rivers 2004.

<sup>8</sup> A concern is that some of these estimates may be contaminated. Suppose party leaders are able to use side payments to force backbenchers to vote for leadership-sponsored proposals that they would otherwise oppose. Such behavior could exacerbate polarization of ideal points and shift the location of uncovered sets calculated from these ideal points. If so, an uncovered set offset toward the majority party would itself be evidence of conditional party government in action, and party influence would again be unobservable. With these concerns in mind, we replicated our analysis using data from Groseclose and Snyder (2000), who estimate two sets of ideal points for the US House: one based on all votes and one based only on lopsided votes (where the winning side received more than 65% of votes cast). Lopsided votes are unlikely to have been the focus of leadership efforts (King and Zeckhauser 2003). Analysis of uncovered sets calculated from lopsided data yields results that are very similar to those presented here, suggesting that contamination is not an issue for our findings.

<sup>4</sup> Regardless of what legislators' utility functions look like, it seems safe to say that they prefer outcomes that are closer to their ideal points to those farther away.

<sup>5</sup> This center of gravity of the majority party's ideal points represents the consensus in the majority party and is used to simplify calculation. Alternatives include the majority median on both dimensions or the majority party uncovered set. All of these measures yield similar results.



- Constant-space ideal points calculated using NOMINATE (Poole and Rosenthal, 1997) for the 81st, 86th, 91st, 96th, 101st, and 106th U.S. House of Representatives.
- Ideal points derived from a linear model for the same House sessions (Groseclose and Snyder 2000).
- Ideal points for the 106th House calculated with a Markov Chain Monte Carlo technique (Jackman, Clinton, and Rivers 2004).
- Ideal points for ten state legislatures from the late 1990's (Aldrich and Battista, 2002) calculated using NOMINATE.

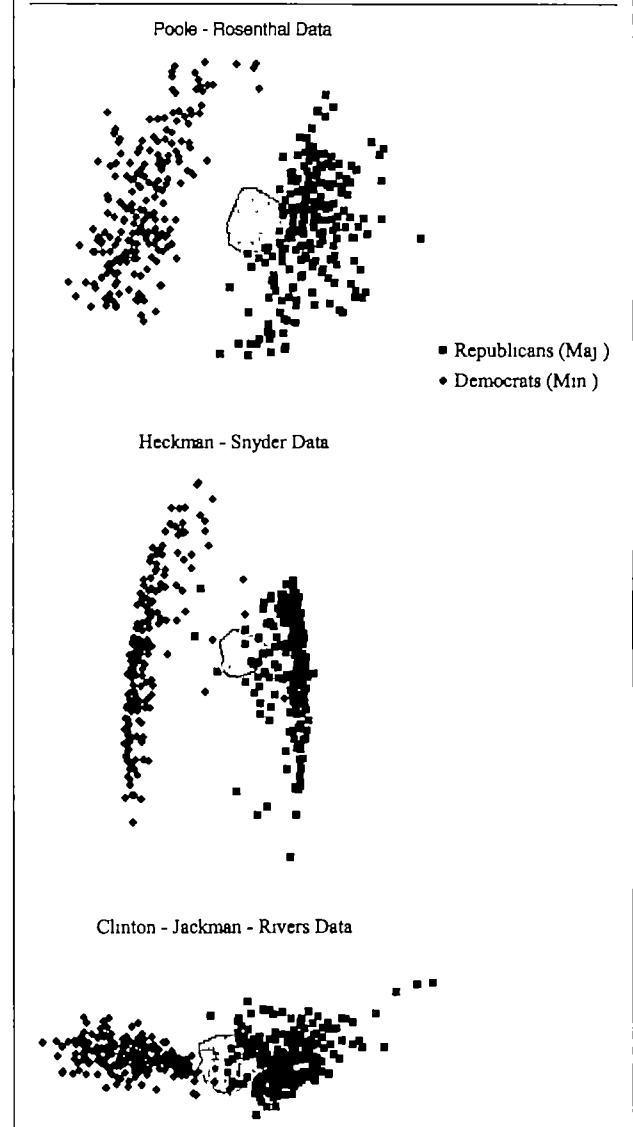
Our analysis strategy is as follows. First, we present uncovered sets for the 106th U.S. House calculated from three different datasets. Although the distribution of ideal points and the size and location of the uncovered set vary across these plots, all show the same pattern: a substantial uncovered set that is closer to the majority-party legislators compared to those in the minority party. Our next step examines Krehbiel's (1999) conjecture about polarization and outcomes by comparing the average distance of uncovered sets to majority- and minority-party legislators. Then, we analyze the potential for agenda setting inside these uncovered sets.

### The Observability of Majority-Party Power

Our analysis provides strong support for the idea that when legislators are polarized by party, the uncovered set is closer to the majority party than to the minority party); thus the measurement of party influence is confounded by the majority party's natural advantage. Figure 3 gives three examples for the 106th U.S. House. The top plot in Figure 3 gives ideal points, and the uncovered set for the 106th House calculated using Poole–Rosenthal ideal points; the middle plot uses Groseclose–Snyder estimates; the bottom uses Jackman–Rivers scores. In all three plots, ideal points for minority-party Democrats are denoted as diamonds and located on the left-hand side, whereas ideal points for majority-party Republicans are squares on the right-hand side. The shaded region is the uncovered set estimated using our grid-search procedure.

The plots show that the different ideal point estimation techniques yield different results, both in terms of the preferences ascribed to each legislator and in terms of the uncovered sets calculated from these ideal points. Such variation may be due to differences in the estimation techniques, or auxiliary assumptions such as the salience of each dimension or the exclusion of certain votes (Jackman, personal communication). Even so, the uncovered sets are similar in two important respects. First, they are not located in the center of the distributions of ideal points. Rather, they are shifted toward the cluster of majority-party ideal points. In addition, the plots reveal that the uncovered sets are fairly substantial in size, occupying about 10% of the Pareto Set. In other words, the intuition of the MVT, that only one outcome is en-

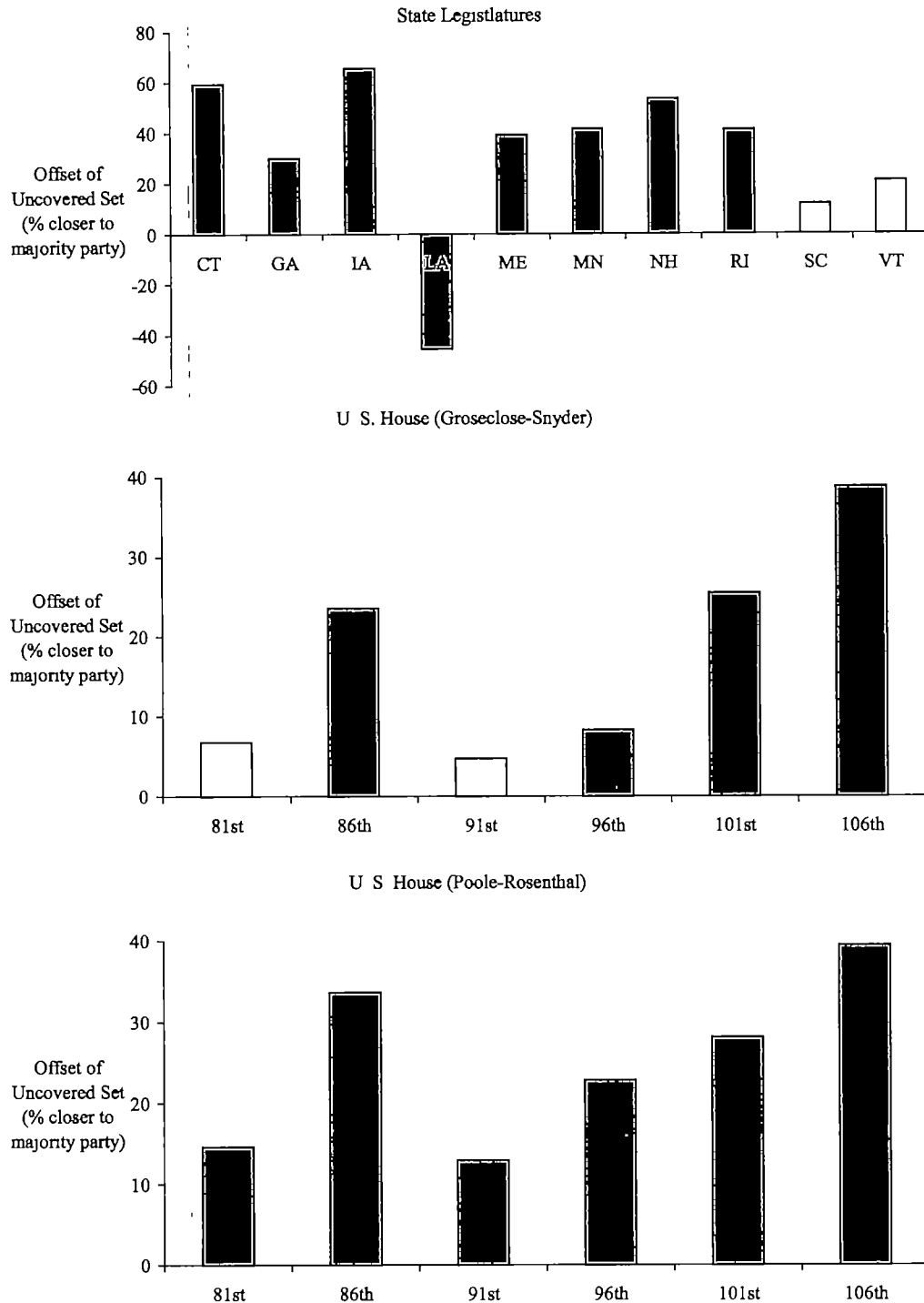
**FIGURE 3. Uncovered Sets for 106th U.S. House**



actable, does not hold in more realistic multidimensional spatial games. This finding is consistent with our hypothesis about the potential gains from agenda setting.

Figure 4 expands our analysis of the location of uncovered sets in real-world legislatures by reporting the relative position of the uncovered set in eleven U.S. state legislatures and six sessions of the U.S. House. As noted earlier, we have two sets of ideal points for five House sessions and three sets for one session. We omit the single data point from Jackman, Clinton, and Rivers, 2004 in this and subsequent figures; however, these data are consistent with those presented here.

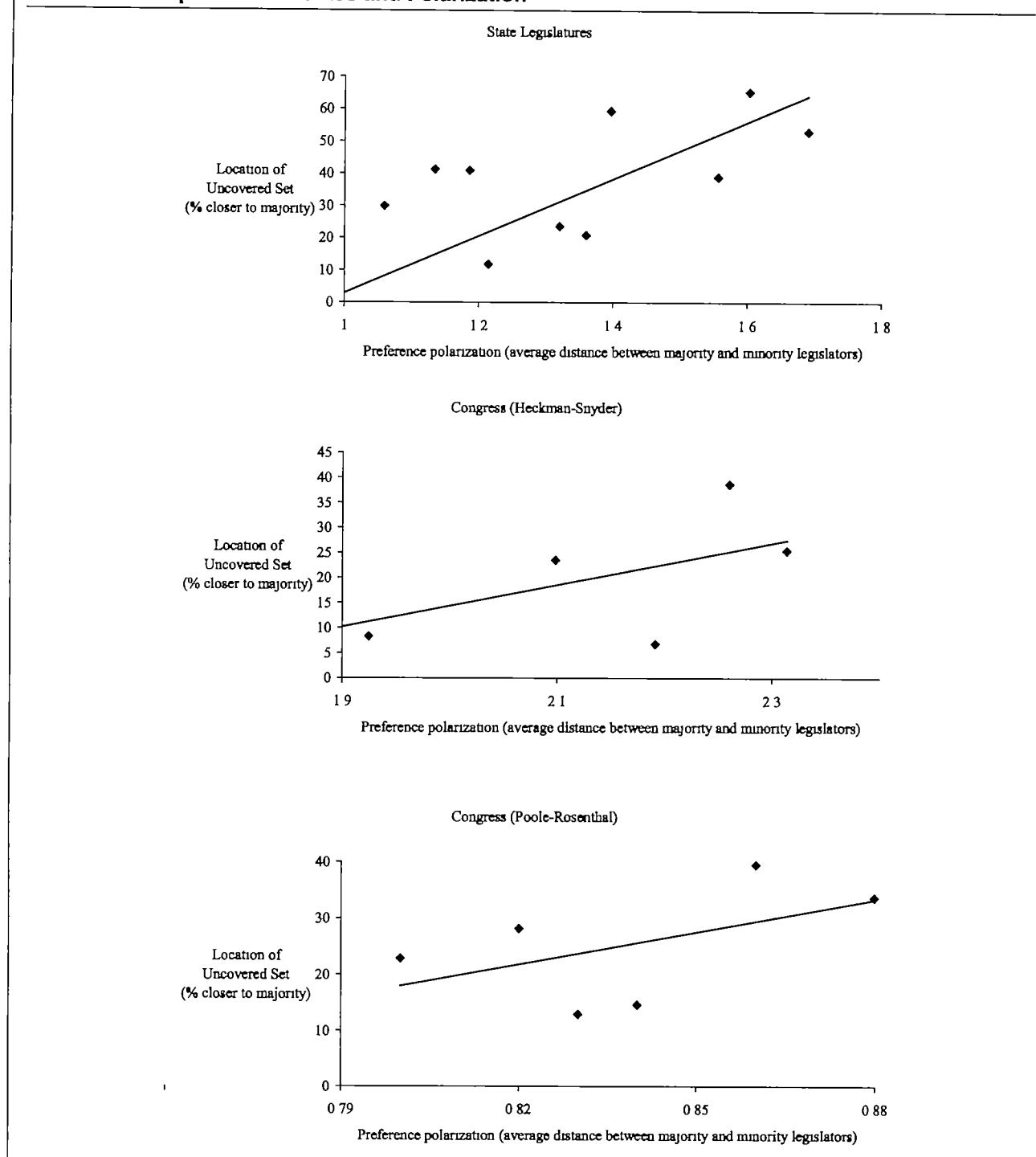
The number reported for each legislature is the ratio of the average distance between majority-party legislators and uncovered set outcomes,  $D_m = \sum_{i \in M, j \in U} D_{ij} / (m \cdot u)$ , and the average distance between minority-party legislators and uncovered set

**FIGURE 4. The Majority Party's Built-In Advantage**

outcomes,  $D_n = \sum_{i \in N, j \in U} D_{ij} / (n \cdot u)$ , expressed as a percentage  $100 \cdot (D_n - D_m) / D_n$ . For clarity, we present plots for different venues (state legislatures vs. the U.S. House) and different methods of calculating ideal points (Heckman-Snyder vs. Poole-Rosenthal). Percentage difference bars that are statistically significant at .05 or better are in dark color; empty bars reflect lower significance levels.

As the figure shows, with one exception, all of the legislatures in our analysis have uncovered sets that are closer on average to majority-party legislators than to legislators in the minority party.<sup>9</sup> Note that in both

<sup>9</sup> The exception is the Louisiana state house, where the majority party has nearly 80% of the seats. The level of polarization in this

**FIGURE 5. Expected Outcomes and Polarization**

congressional plots, the offset of the uncovered set increases over time, suggesting that at least part of the apparent increase in party influence in the postwar House is due to the polarization of the party caucuses.

legislature is also the lowest of all of the cases in our analysis. We conjecture that in one-party legislatures, the logic of cross-party competition breaks down and conditional party government involves a faction within the majority, or a cross-party coalition, a speculation consistent with Jenkins and Weidenmier's (1999) analysis of the early-1800s U.S. Congress

Figure 5 extends the analysis, showing that the position of the uncovered set relative to the majority and minority parties is influenced by polarization: as the average distance between majority- and minority-party legislators increases, the uncovered set moves relatively closer to the majority party.

These findings about the location of uncovered sets in real-world legislatures suggest that concerns over the observability of party influence are well founded. Regardless of whether preferences are specified using

one dimension or two, when legislators are polarized by party, the set of feasible outcomes favors is closer to the majority party. Thus, an observer who considered actual outcomes relative to legislators' ideal points without examining the uncovered set would conclude that the location of these outcomes indicated a majority-party cartel at work, when in fact party leaders might be inactive or powerless. In this sense, our analysis provides a partial confirmation of Krehbiel's (1999) conjecture concerning the observability of party influence.

### Agenda-Setting and Majority-Party Influence

The fact that polarization gives the majority party a built-in advantage in terms of outcomes does not imply that conditional party government is impossible. As noted earlier, when the uncovered set contains many outcomes, majority party leaders can pick one and formulate an agenda that yields this outcome as the result of majority-rule voting. Earlier, we described distance measures that characterized this potential. Figure 6 reports these measures for the three datasets in our analysis—again, we separate results by estimation method and venue.

The results highlight the potential for agenda-setting: in all cases, the uncovered set outcome that is closest to the majority-party consensus is noticeably closer than the average uncovered set outcome. For example, in the case of the 106th House using the Groseclose–Snyder data, the average distance of uncovered set outcomes to the caucus center of gravity point,  $\rho_m$ , is about 25% of the  $x$ -axis range (note the uncovered set for this session is shown in Figure 3). However, by using an agenda strategy that yields the closest possible uncovered set outcome, majority party leaders can cut this distance in half: the distance between the majority center of gravity and this outcome is only 13% of the  $x$ -axis range. Additional analysis—omitted here, but available on request—shows that the difference in the two distances is higher given higher levels of polarization.

In substantive terms, these results show that in real-world legislatures, by choosing an appropriate agenda, party leaders can move legislative outcomes considerably closer to the preferences of their caucus compared to the expected results given leader inaction. Because these efforts involving movement within the uncovered set—that is, selecting one enactable outcome and devising an agenda that yields it—this potential for majority-party influence exists at the margin of whatever inherent advantages are conveyed to the majority by the location of the uncovered set.

The charts also show that the potential gains from agenda setting vary across legislatures. For example, among the state legislatures, some states (e.g., Maine) have uncovered sets that are close to the majority caucus to begin with (low  $C$ ), whereas in others, the uncovered set lies farther away (e.g., Connecticut, where  $C$  is high). Similarly, in some states, agenda setting provides relatively substantial gains over the situation where leaders are inactive (e.g., Vermont; note the difference

between  $C$  and  $C_x$ ), whereas in other states the gains are more modest (e.g., South Carolina, where  $C$  and  $C_x$  are similar).

These results suggest that the value of conditional party government is not the same for all legislatures. Polarization may be a necessary condition for conditional party government to operate, but it is not sufficient. Depending on the distribution of legislators' preferences, both the level of disagreement between the parties and the level of agreement within the majority party, legislators in the majority caucus may decide that the range of outcomes that are possible given an inactive leadership are sufficiently good that the cost of empowering leaders outweighs the gains.

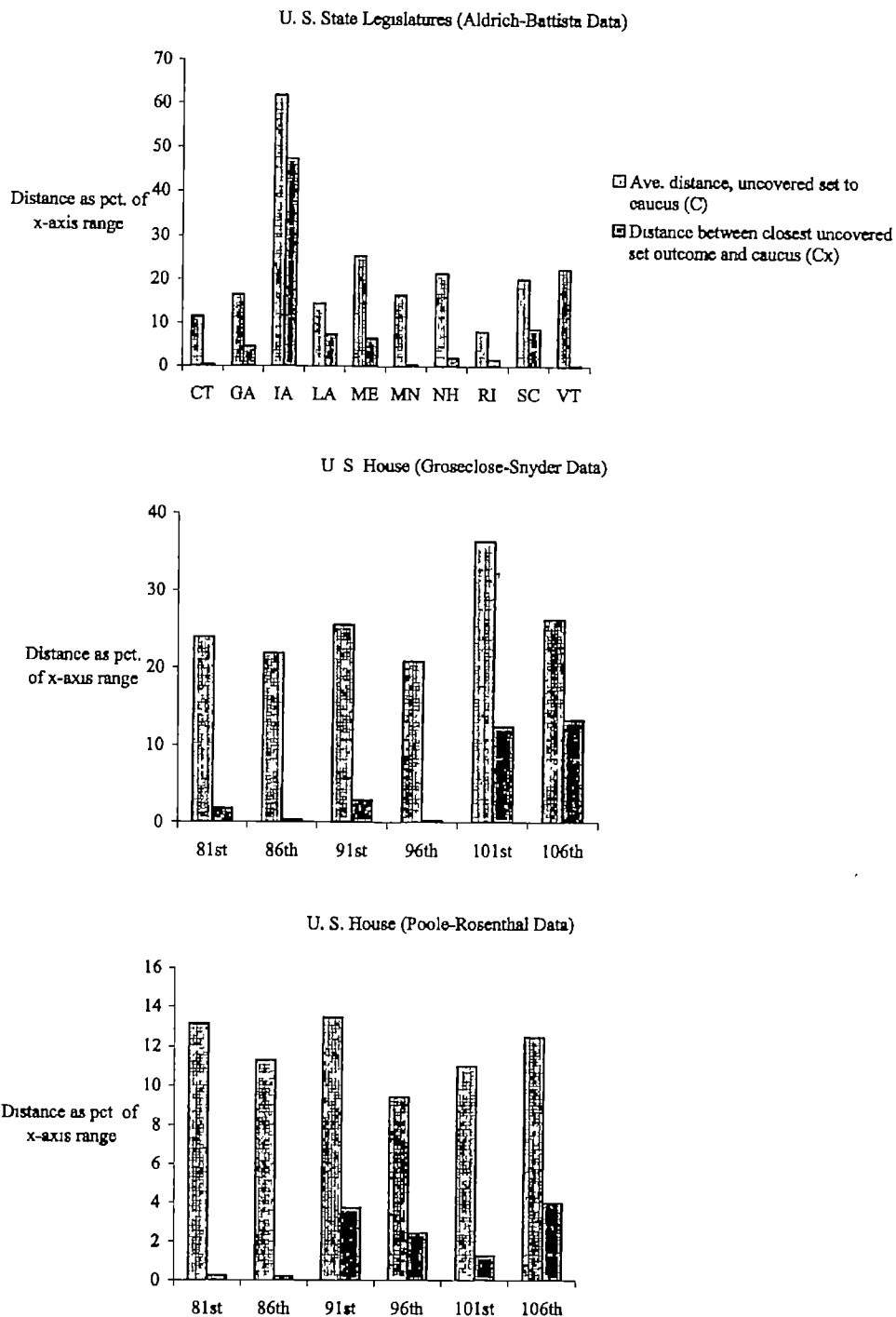
More generally, the data in Figure 6 only describe the potential gains from agenda setting. There is no assurance that the majority caucus will agree on which outcome to implement or that party leaders will respond to a caucus mandate by devising appropriate agenda strategies. What these results establish is that in a realistic model of the legislative process, the potential exists for the majority party to use its control over legislative procedures to make real improvements in legislative outcomes—improvements that occur at the margin of whatever advantages are conveyed to the majority party by polarization.

### DISCUSSION

This paper aims at resolving the debate over majority party influence in contemporary American legislatures. Our use of new analytic techniques enabled us to begin with a better model of legislative proceedings—to abandon simple one-dimensional spatial models in favor of more realistic two-dimensional versions. Our conclusions are based on the analysis of real-world data rather than on arguments about the relative merits of different theoretic assumptions.

Our analysis confirms that when legislators' preferences are polarized, outcomes will generally be closer to the majority party's wishes, even if the majority party leadership does nothing to influence the process by which proposals are offered, amended, and voted on. Put another way, even in a multidimensional model of legislative proceedings, a single-minded focus on outcomes and preferences will tend to overstate the majority party leadership's influence over the legislative process.

However, our analysis also shows that at the margin of the majority party's natural advantage in a polarized legislature, agenda-setting remains an efficacious strategy. Previous analyses of conditional party government, which framed the legislative process in terms of a single policy dimension, assumed this possibility away, for in these settings, party leaders' only option is to accede to the preferences of the median floor legislator. Our work analyzes party influence using a two-dimensional framework, exploiting a new technique for determining enactable outcomes, or the uncovered set, given real-world preference data.

**FIGURE 6. The Potential for Majority-Party Influence via Movement within the Uncovered Set**

We find that for all of the legislatures analyzed here, the potential exists for majority party leaders to use agenda-setting strategies to move outcomes closer to the preferences of their caucus. The magnitude of these potential gains varies across legislatures but always exists to some degree.

The limits of our findings bear emphasis. As noted earlier, the potential for party influence varies across legislatures, most notably with the distribution of

legislators' preferences, which in turn shape the size and location of the uncovered set. Moreover, our analysis has only considered the potential for party influence—we have not examined whether majority-party leaders actually implement, or try to implement, agenda setting strategies. Finally, our analysis says nothing about other mechanisms of majority-party influence, such as strategies involving committee jurisdictions or assignments.

Notwithstanding these caveats, this paper shows that conditional party government is more than a description of how real-world legislatures appear to operate. Rather, the theory's expectations of how the majority party shapes legislative proceedings, as well as its claims about the potential gains from these strategies, are consistent with over two generations of work on spatial models of legislative action and supported by empirical analysis. With this rationale in hand, the next step is to begin the systematic testing of hypotheses about majority-party influence.

## REFERENCES

- Aldrich, John, and David Rohde. 1998. "Theories of Party in the Legislature and the Transition to Republican Rule in the House," *Political Science Quarterly* 112: 112–35.
- Aldrich, John H., and David W. Rohde. 2001. "The Logic of Conditional Party Government: Revisiting the Electoral Connection." In *Congress Reconsidered*, Lawrence Dodd and Bruce Oppenheimer, eds. Washington: CQ Press.
- Aldrich, John H. 1995. *Why Parties? The Origin and Transformation of Political Parties in America*. Chicago: University of Chicago Press.
- Aldrich, John H., and James S. Battista. 2000. "Conditional Party Government in the States" *American Journal of Political Science* 46: 235–56.
- Bianco, William T., Ivan Jeliaskov, and Itai Sened. 2004a. "The Uncovered Set and the Limits of Majority Rule." *Political Analysis* 12: 256–76.
- Bianco, William T., Michael S. Lynch, Gary J. Miller, and Itai Sened. 2004b. "A Theory Waiting to Be Discovered and Used: A Reanalysis of Canonical Experiments on Majority Rule Decision-Making." Unpublished manuscript.
- Bianco, William, James Holloway, Michael Lynch, Gary Miller, and Itai Sened. 2004c. "Constrained Instability: Experiments on the Robustness of the Uncovered Set." Unpublished manuscript, presented at the 2004 APSA meetings, Chicago, IL.
- Cox, Gary W. 1987. "The Uncovered Set and the Core" *American Journal of Political Science* 31/2: 408–22.
- Cox, Gary W., and Mathew D. McCubbins. 1993. *Legislative Leviathan. Party Government in the House*. Berkeley: University of California Press.
- Cox, Gary W., and Mathew D. McCubbins. 2003. *Legislative Leviathan Revisited*. Unpublished manuscript.
- Clinton, Joshua, Simon, Jackman, and Doug Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2): 355–70.
- Cox, Gary W., and Keith T. Poole. 2002. "On Measuring Partisanship in Roll-Call Voting. The U.S. House of Representatives, 1877–1999." *American Journal of Political Science* 46: 477–89.
- Groseclose, Tim, and James M. Snyder, Jr. 2000. "Estimating Party Influence in Congressional Roll-Call Voting." *American Journal of Political Science* 44 (2): 193–211.
- Groseclose, Tim, and James M. Snyder, Jr. 2001. "Explaining Party Influence in Roll Call Voting: Regression Coefficients vs. Classification Success" *American Political Science Review* 95: 689–98.
- Jenkins, Jeffrey, and Marc Weidenmier. 1999. "Ideology, Economic Interests, and Congressional Roll-Call Voting. Partisan Instability and Bank of the United States Legislation, 1811–1816." *Public Choice* 100: 225–43.
- King, David C., and Richard L. Zeckhauser. 2003. "Congressional Vote Options." *Legislative Studies Quarterly*, 28 (August): 387–411.
- Krehbiel, Keith. 1999. "Paradoxes of Parties in Congress" *Legislative Studies Quarterly* 24: 31–64.
- Krehbiel, Keith. 2000. "Party Discipline and Measures of Partisanship." *American Journal of Political Science* 44: 212–27.
- Londregan, John. 1999. "Estimating Legislators' Preferred Points." *Political Analysis* 8: 35–56.
- McCarthy, Nolan, Keith Poole, and Howard Rosenthal. 2001. "The Hunt for Party Discipline in Congress" *American Political Science Review* 95: 673–88.
- McKelvey, Richard D. 1986. "Covering Dominance and Institution Free Properties of Social Choice," *American Journal of Political Science* 30: 283–314.
- Miller, Nicholas. 1980. "A New Solution Set for Tournament and Majority Voting." *American Journal of Political Science* 24: 68–96.
- Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll-Call Voting*. New York: Oxford University Press.
- Shepsle, Kenneth A. 1979. "Institutional Arrangements and Equilibria in Multidimensional Voting Models." *American Journal of Political Science*, 23: 27–59.
- Shepsle, Kenneth, and Barry R. Weingast. 1984. "Uncovered Sets and Sophisticated Voting Outcomes with Implications for Agenda Institutions." *American Journal of Political Science* 28: 49–74.
- Shepsle, Kenneth A., and Barry R. Weingast. 1994. "Positive Theories of Congressional Institutions," *Legislative Studies Quarterly* 19: 149–81.



# Racism Revised: Courts, Labor Law, and the Institutional Construction of Racial Animus

PAUL FRYMER *University of California, Santa Cruz*

*How should we understand and explain individual acts of racism? Despite extensive debate about the broader place and importance of racism in America, there is surprisingly little theoretical or empirical analysis of what leads individuals to commit racist acts. In contrast to most political scientists who understand racism as an individual psychological attitude—an irrational prejudice—I argue that individual manifestations of racism are the result of a complex set of factors, and that latent psychology is less helpful to understanding them than are the maneuverings and behavior of strategic actors following rules and incentives provided by institutions. We need to examine the ways in which institutions encourage racist acts by motivating people to behave in a racist manner or behave in a manner that motivates others to do so. To further explore and compare institutional and individual-psychological approaches to understanding racism, I examine manifestations of racism in labor union elections. I analyze and contrast more than 150 cases in which the National Labor Relations Board and U.S. federal appellate courts formally responded to reported violations of racism in a union election. The principles of this approach can easily be applied to other contexts and suggests that racism in society is less intractable and innate than malleable and politically determined.*

How should we understand and explain individual acts of racism? Despite extensive debate about the broader place and importance of racism in America, there is surprisingly little theoretical or empirical analysis of what leads individuals to commit racist acts. Although scholars differ on racism's particularities, extensiveness, and significance, most political scientists are in agreement that racism is at its root an individual psychological attitude, an irrational prejudice, that stems from feelings of resentment and animus toward others or from a desire to create group status hierarchies (e.g., Bobo 1988; Kinder and Sanders 1996; Sears 1988; Sidanius and Pratto 1999; Sniderman and Piazza 1993). Only a few studies explore what leads people to act on their attitudes and those that do tend to similarly emphasize individual factors such as the racist actor's feelings of resentment, insecurity, and anxiety (e.g., Gourevitch 1999; Green et al. 1998; Olzak 1992). Because scholars see racism as an irrational prejudice, they tend not to examine the role that rules, institutions, and politics play in determining why some individuals act on their racist attitudes and why others do not. Certainly, in comparison to the wealth of studies documenting the importance of rules and institutions in guiding the behavior of individual legislators, party elites, judges, and public administrators, there has been little examination of why certain individuals commit racist acts.

This model of racism is not confined to public opinion scholars. From de Tocqueville (2001) to Du Bois (1935) to Myrdal (1944) to Roediger (1991), race theorists have consistently conceived of racism as both an individual and psychological phenomenon. Racism is treated as antithetical to ideological traditions of tolerance and equality and is attributed to people wanting to maintain emotionally based hierarchies. Even those who highlight the relationship between racism and power tend to fall back on definitions that see it as a "deformity of rationality" (e.g., Appiah 1990, 8; Rogin 1988; Takaki 2000). For instance, although Albert Memmi (1999, 38, 27) argues that the "machinery of racism" enables elites to exercise power and privilege, he claims that racism arises from an individual's "mistrust, if not repulsion and fear" of something—or someone—who is different, "like an unfamiliar plant growing by the side of the road, whose odor itself may be noxious." Rogers Smith (1997, 38) sees racism as intrinsic to elite efforts at state building; at the same time he argues that racism both derives and sustains itself from individual resentments and the desire, particularly among those less powerful, "to feel part of a larger, more enduring whole of intrinsic worth." Similarly, many critical race theorists who claim that law and institutions enable and legitimate racist activity maintain that racism stems from individual-driven irrational behavior that is unconscious or an "attenuated" psychological predisposition (e.g., Haney López 2000, 1730; Krieger 1995; Lawrence 1987). At base, then, all of these analyses believe individuals buy into the appeal of racism because of psychological needs, not because they are motivated by broader institutional dynamics.

Viewing racism through the lens of psychology is not so much wrong as incomplete. By understanding racial conflict as irrational acts conducted outside the confines of political actors and institutions, these works depoliticize racist activity and ignore important dynamics of power and incentives that shape individual behavior.

Paul Frymer is Associate Professor, Department of Politics, University of California at Santa Cruz, Santa Cruz, CA 95064, pfrymer@ucsc.edu.

Thanks go to Amy Bridges, Michael Brown, David Canon, Kerstin Carlson, John Coleman, Zolt Hajnal, Kim Johnson, Tom Kim, Ari Krantz, Cathie Lee, Michael McCann, Eileen McDonagh, Tali Mendelberg, Shehzad Nadeem, Corey Robin, Mark Sawyer, Pam Singh, John Skrentny, Rogers Smith, Julie Suk, Rick Valelly, Dorian Warren, and Lee Sigelman. The Law and Public Affairs Program at Princeton University and the University of California's Institute for Labor and Employment provided generous financial assistance.

Racism, writes Adolph Reed (2000), "is not an affliction. . . . Nor is it a thing that can act on its own; it exists only as it is reproduced in specific social arrangements in specific societies under historically specific conditions of law, state, and class power." Racist manifestations by individuals are the result of a complex set of factors, and latent psychology, I argue, is less helpful for understanding it than are the maneuverings and behavior of strategic actors following rules and incentives provided by institutions. Moreover, racism is not politically problematic simply because some or even many individuals hold racist attitudes; it becomes problematic when institutional dynamics legitimate and promote racist behavior in a concentrated and systematic manner. As such, we need to examine the ways in which institutions encourage racist acts by providing rules and procedures that motivate people to behave in a racist manner or behave in a manner that motivates others to do so.

The institutional analysis of racism I put forth in this paper draws from and expands on the work of a multidisciplinary group of race scholars. Many of these scholars have examined how racial cleavages intersect with institutional dynamics, leading to racism's continuing importance in America even as societal attitudes seemingly change (e.g., Bonilla-Silva 1996; Frymer 1999; Hochschild 1984; Lieberman 1998). Others have focused on the role of state elites in configuring racial and racist understandings through bureaucracies and political institutions (e.g., Brown et al. 2003; Katznelson 1973; King 1997; Marx 1998; Skrentny 2002; Walton 1997) and through the dissemination of ideologies that either promote racial hierarchies or attempt to reconcile public aspirations for freedom with widespread racial inequality (e.g., Du Bois [1903] 1999; Fields 1982; King and Smith 2005; Smith 1997). Claire Kim (2000, 9) argues for a notion of "racial power" that is not "something that an individual or group exercises directly and intentionally over another individual or group but rather as a systemic property, permeating, circulating throughout and continuously constituting society." Still others have examined the confluence between the state and the market, arguing in different ways that racism is related to class hegemony (e.g., Bobo 1988; Cox 1948; Goldfield 1997; Reed 2002). All of these works intersect with and have been influenced by the scholarship of those who contend that race is continually being formed and re-formed in the context of political struggle (e.g., Gilroy 1984; Hall 2000; Kim 1999; Omi and Winant 1994). Race and racism can take many forms—as Anthony Appiah (1990) argues, there are distinct "racisms"—that are constructed by elites who fight over the meaning of the terms in an effort to maintain power and promote differently racialized agendas.

This previous work almost exclusively focuses on race and racism at a theoretical and/or macro level (but see Haney López 2000; Kim 2000). In doing so, it has offered compelling arguments for how racism is a part of our national identity and embedded in and shaped by "the State," but it has not examined how concrete institutional dynamics work to motivate indi-

vidual behavior. Even studies of "institutional racism" emphasize less how institutions motivate individuals than simply making the important point that racism is located within places of power (e.g., Carmichael and Hamilton 1967; Knowles and Prewitt 1969). By contrast, the institutional understanding of individual racism I put forth emphasizes the following features, all of which are fairly common to institutional studies of politics but have not been applied to understandings of individual racism. First, institutions do not merely provide avenues for racist actors to operate, but can independently encourage racist acts by influencing individual preferences and rewarding certain types of behavior over others—they are, as March and Olsen (1984, 738) argue, "more than simple mirrors of social forces." Second, institutions enhance the power of those actors who are well situated within them, providing these actors with coercive power, the ability to set the agenda, and the ability to anticipate and in turn shape the behavior of those with whom they interact (e.g., Gaventa 1980; Lowi 1969). Third, not all forms of political and individual behavior result in the same opportunities and outcomes; collective action problems in particular disadvantage some interests and forms of political action while promoting others (Mansbridge 1986; Olson 1971). To understand manifestations of individual racism, we must recognize that institutional structure and organizational dynamics influence whether racist actors express themselves or whether they remain silent. Fourth, the use of race and racism is not, by itself, politically problematic, nor are all racist expressions equal in significance. Only by placing the manifestations of individual racism in a broader context can we understand how the act acquires importance and meaning (Balibar 1992; Omi and Winant 1994; Said 1990). By deemphasizing the importance of individual prejudice—although by no means denying its existence—as a determinative feature of racist manifestations, I wish to locate the act within the context of institutional combat. Such an understanding, in turn, sees racism in society less as intractable and innate than as something malleable and politically determined.

To further explore and compare institutional and individual approaches, I examine manifestations of racism in labor union elections. I analyze more than 150 cases in which the National Labor Relations Board (NLRB) and federal appellate courts have formally responded to reported instances of racism during a union election. Racist acts in union elections are considered an unfair labor practice under national labor law and either the NLRB or a federal appellate court can overturn an election's outcome if it finds that those acts unduly influenced the voters' decisions. The holdings of the NLRB and federal courts, as well as extensive detail of the facts and context of the racist acts, are publicly available and the data set that I have compiled is the universe of reported cases between 1935 (the year that the National Labor Relations Act was passed) and 2000. The data set is unique and relevant for many reasons. Most important, the thick descriptions of each case provide an opportunity to analyze racism within a

broader legal and institutional context of workers and employers strategically vying for power over the workplace. Moreover, it offers an excellent opportunity to compare the theoretical leverage provided by individual and institutional models of racism: as we shall see, federal courts typically respond to reports of racism in a manner parallel to the individual-psychological model endorsed by most political scientists whereas the NLRB consistently treats the same factual events as engrained in, and as a product of, institutions. Analyzing how two different legal bodies come to often entirely different interpretations of the same incident allows us to see both the advantages and disadvantages of each theoretical approach and to highlight the assumptions that underlie both. The data set, it should be noted, is also limited in important ways. It relies entirely on published cases and cannot include the untold number of situations that either went unreported or did not warrant a response in the Board's estimation. The available data illustrate how contrasting theories of racism explain individual acts but do not provide a "test" of the models in any way nor, because of inherent problems with sampling bias, an empirically conclusive comparison of Board-court behavior. The goal of this paper, then, is at the level of theory: to illustrate both the limits of an individual approach to understanding racism and the theoretical contributions of an institutional approach.

## TWO VIEWS OF RACISM: ANTI-DISCRIMINATION AND LABOR LAW

American employers, workers, and labor unions have a long history of participating in workplace racism, producing racially segregated and unequal workforces and unions (see, e.g., Nelson 2001; Roediger 1991). Federal labor law in the first few decades of the twentieth century gave unions and employers the opportunity to sign collective bargaining agreements that provided for white-only workforces, leading to the effective removal of African Americans and other racial minorities from whole spheres of employment (Arnesen 2001; King 1997). By the mid-twentieth century a significant number of national and local unions participated in systematic and widespread discrimination, particularly against African American workers, by denying them employment and union representation, committing unfair labor practices, and participating in explicit and often violent racial conflict (Gould 1977; Hill 1985).<sup>1</sup> Both the federal courts and the NLRB have actively intervened, but the two institutions have consistently responded to union racism in fundamentally different ways. Courts focus on the wrongdoing of the actor and view racist acts as outside the confines of rational politics, and the NLRB, although finding racism virulent and wrong, has tended to treat it as part and parcel of a broader set of institutional and political dynamics.

Federal courts have confronted union racism primarily through Title VII of the 1964 Civil Rights Act and the Equal Protection clause of the Fourteenth Amendment. Both legal instruments have been used in different ways at different times by judges of different political persuasions and scholars have thoroughly argued that multiple traditions of antidiscrimination law have had moments of resonance (e.g., Balkin and Siegel 2003; Forbath 1999; Kersch 2004). Despite this variety, one assumption remains quite persistent, particularly in the post-Civil Rights era: in a democratic and liberal society that values individual tolerance, racism is wrong, irrational, and should always be outside of politics and law. Racism is "obviously irrelevant and invidious," declared the Supreme Court in perhaps its most famous union discrimination case (*Steele v. Louisville & N.R. Co.*, 323 U.S. 192, 203 [1944]). "A racial classification, regardless of motivation, is presumptively invalid" (*Personnel Administrator of Massachusetts v. Feeney*, 442 U.S. 256, 272 [1979]) and places a "brand upon" those who are its targets (*Strauder v. West Virginia*, 100 U.S. 303, 308 [1880]). The Constitution, after all, is "color-blind" (*Plessy v. Ferguson*, 163 U.S. 537, 559 [1896], Harlan dissenting), and race and racism should have no place in political dialogue and involvement—indeed, the racial classification itself constitutes a *prima facie* indicator of discrimination and prejudice (e.g., *Anderson v. Martin*, 375 U.S. 399 [1964]; *Loving v. Virginia*, 388 U.S. 1 [1967]; *Shaw v. Reno*, 509 U.S. 630 [1993]). In contrast to some forms of government classifications on the basis of sex, sexuality, disability, and age that the Supreme Court has argued have at least the potential of being rational and legitimate considerations, racial classifications are given the highest level of scrutiny. The Court strongly suspects racial classifications are motivated by invidious and irrational goals and assumptions and will only allow the classification if the government can provide a "compelling" reason (Brest 1976; Post 2000). When a broader political or social context has been introduced as a rationale for the government's racial classification, the Supreme Court has been fairly consistent in rejecting the broader context unless specific individual-level racism can be proven (e.g., *McKleskey v. Kemp*, 481 U.S. 279 [1977]; *Milliken v. Bradley*, 418 U.S. 717 [1974]; *Wygant v. Jackson Board of Educ.*, 476 U.S. 267 [1986]).

The Court's interpretation of Title VII has followed similar assumptions to the aforementioned Equal Protection cases. In Title VII cases, the issue is less about a racial classification than about the motivations behind a purported incident of discrimination in the workplace, usually by an employer. A typical case involves a judge attempting to locate (often with the help of psychologists) the precise moment when an individual is specifically and directly motivated by racism (or sexism or other forms of animus), and to separate this moment from other moments when the individual is presumably motivated by rational pursuits (e.g., *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502 [1993]; *Price Waterhouse v. Hopkins*, 490 U.S. 228 [1989]; *Alexander v. Sandoval*, 532 U.S. 275 [2001]). Sometimes, this precise moment and motive is difficult to prove and courts have used a

<sup>1</sup> In claiming that unions had significant race problems, I am dramatically simplifying a complex story, overlooking those unions whose leaders and members worked actively for civil rights (see, e.g., Draper 1994; Kelley 1990).

variety of measures to make it easier for a lawsuit to go forth without a “smoking gun” to find whether animus was the pretext for the actor’s otherwise seemingly rational decision, whether by making it relatively easy for a plaintiff to establish a *prima facie* (*McDonnell Douglas Corp. v. Green*, 411 U.S. 792 [1973]) or by allowing for statistical evidence to show a “disparate impact” (*Griggs v. Duke Power Co.*, 401 U.S. 424 [1971]) and a “pattern and practice” of discrimination absent a finding of individual intent (e.g., *Teamsters v. United States*, 431 U.S. 324 [1975]). Critical to current antidiscrimination law, however, is the sense that an individual or group of individuals is responsible for the racist act, that they must be punished, and that such behavior must be removed from the sphere of rational decision making (see Freeman 1978; Krieger 1995). Moreover, similar to the assumptions behind the Supreme Court’s Equal Protection decisions, Title VII law declares that race can never be a factor in rational decision making, even when the law explicitly makes an exception for other forms of discrimination on grounds that gender or disability can, under certain conditions, be objective considerations for employment.<sup>2</sup>

Premised on an entirely different understanding of individuals and power, labor law provides a sharp and interesting contrast to the courts’ approach to racism. Congress passed the National Labor Relations Act in 1935 to promote peaceful resolution of workplace conflicts by giving unions the opportunity to engage in collective bargaining with employers. In passing the Act, legislators recognized that workers and management were fundamentally at odds, and that the NLRB was given regulating power to negotiate contracts and bargaining agreements between the two sides, all the while recognizing the potential dangers of discrimination and intimidation by both employers and unions during the context of union activity. Unlike most government institutions that claim to operate in an environment of more or less equally powerful actors, and unlike federal courts which are dominated by an anti-classification model of discrimination that looks only for the mere mention of race, the NLRA uniquely recognizes the fundamentally unequal power relations of the workplace and the “relative weakness of the isolated wage earner” (Senate Report No. 573, 74th Cong., 1st Sess., 1935, 3). Interpreting the Act in *NLRB v. Jones & Laughlin Steel Corp.* (301 U.S. 1, 33 [1937]), the Supreme Court acknowledged “that a single employee was helpless in dealing with an employer; that he was dependent ordinarily on his daily wage for the maintenance of himself and family . . . that a union was essential to give laborers opportunity to deal on an equality with their employer.” More than three decades later, the Court reiterated in *NLRB v. Gissel Packing Co.* (395 U.S. 575, 617 [1969]) that any rights of the employer to promote its position to the workers during a union drive must be balanced by “the economic

dependence of the employees on their employers, and the necessary tendency of the former, because of that relationship, to pick up intended implications of the latter that might be more readily dismissed by a more disinterested ear.” Although the NLRB is an agency that continually changes based on the preferences of its politically appointed members, its statutory obligations and regulatory nature have led it to consistently scrutinize individual actions in the context of NLRA rules and the broader goals of the Act.

The NLRB’s handling of union elections typifies this approach to the individual actor. Labor unions are most commonly formed through employee elections. These elections are unlike other democratic elections in the United States in that the free speech of all the participants involved—the workers who are voting, the union organizers, and the employers—is severely restricted. Labor law gives voters in union elections “the opportunity of exercising a reasoned, untrammelled choice for or against labor organizations seeking representation rights” (*Sewell Manufacturing*, 138 NLRB 66 [1962]). The NLRB has consistently worried that workers will treat speech during elections as a threat that can be carried out, usually by employers because they control hiring and firing, pay, benefits, and promotions. Thus, the Board scrutinizes acts and speech during the election campaign that might be conceived as a threat—explicit or implicit—and, as a result, lead workers to vote in a manner that does not reflect their true preferences. Over the years, the NLRB has devised a number of rules governing the speech and conduct of the relevant actors during union elections, which are arranged and directly supervised by regional boards around the country (for an overview of NLRB election policy, see Becker 1993; Getman et al. 1976). Employers, for instance, are allowed to state their opposition to unions but are not allowed to threaten reprisals or promise benefits. No unlawful firings are allowed, no threats of plant closure or loss of jobs or denial of future benefits can be made, no promises of specific future benefits may be suggested, nor are bribes by the employer or by the union to employees allowed. The employer is not allowed to interrogate employees about their union sympathies, nor is it allowed to improperly survey union activities. The employer may not make campaign speeches to the employees within 24 hours of the election. At the same time, the union cannot pass out literature and campaign on workplace grounds (though the employer can), cannot participate in “captive audience speeches” held by the employer (unless the employer chooses otherwise), and, similar to the employer, cannot threaten the workers in any way. If a violation of these rules occurs during the election period, the Board has the power to overturn election results or, in more extreme cases, to issue injunctions and/or bargaining orders that force the employer to sign a contract with the union even if the union has lost the election.

Racism by one of the parties during the campaign is a violation that can lead the Board to overturn an election. Unlike the federal court decisions under antidiscrimination law, racism in labor law is regulated

<sup>2</sup> Under Title VII law, there is a “bona-fide occupational qualification” exception that enables an employer to discriminate on the basis of gender and age, but not race, if the discrimination is deemed essential to the job criteria

for its potentially damaging political consequence, not because it is considered reprehensible and unacceptable in any context. The Board approaches racist acts through Section 8(a) of the NLRA, which prohibits threats to workers made in the context of union election campaigns. The Board has feared that, in the effort to dissuade employees from voting in favor of the union, the employer may resort to a variety of tactics designed to inflame racial prejudice among employees and, as a result, convince workers that voting for a union will hurt their workplace environment. Alternatively, the union may attempt to arouse feelings of racial pride in employees either by attempting to create an exclusive racial hierarchy among one group of workers at the expense of another, or to attack the employer on racial grounds, or to convince the employees that they need a union for protection from a racist employer. The individual worker is clearly the object of these appeals and the fact that workers respond to these tactics is no doubt the result of at least implicit support—whether psychological or ideological—for such ideas. But the ideas in the abstract are not what the Board sees as problematic, rather the Board focuses on two things; which side prompted the racist act and if the act was consequential. And here, unlike the view of courts and much of political science, the causation lies not in individuals wishing to benefit from an emotive or psychological wage, but from institutional dynamics that promote this type of behavior among employers and union leaders to obtain their broader goals.

The Board perceives racism, as the following pages show, as a politically rational strategy employed by both sides during union elections. One side wants to use race and racism to divide workers, the other side to unify workers. The institutional context in which union elections are waged allows race-based strategies to be used in many different ways, by many different groups, and often in a less than explicit manner. The Board's response is multifaceted. First, it has defined racism in the context of whether it is threatening or not. If racism is seen as economically or politically coercive, it is regulated; if not, it is usually allowed. Second, there is an institutional dimension that is relevant, unbalanced, and an important influence on individual behavior. Employers must bargain with the union once it is elected and, thus, are not allowed to intimidate or coerce during the campaign. At the same time, the employer retains certain powers, most notably the power to hire new workers, to shut down or move the company out of the United States, to play hardball in union negotiations, and to hire replacement workers if a strike were to ensue. The Board is highly attuned to an employer's power to set agendas and manipulate institutional rules and disallows racism to the extent that it contributes to that power, particularly when its introduction to an election campaign seems a product of employer efforts to defeat the union. Third, the explicit assumption of labor law is that workers cannot succeed without collective action, and so the Board tends to discount individual preferences in favor of majority rule. Because the Board focuses more on majorities than individuals, it usually differentiates

between leaders who are deemed accountable to their constituencies for their actions and individuals who are not. This means that the Board examines the identity of the perpetrator of the racist act—whether he or she is an authorized leader or merely an individual—to determine whether the act is politically significant and worth regulating.

## UNION ELECTIONS AND LABOR REGULATIONS AGAINST RACISM

### The NLRB Cases: Agenda Setting, Power, and Collective Action

The three aforementioned institutional dynamics broadly explain the Board's handling and understanding of union racism. In this section, I examine more than six decades of decisions in which the Board confronts racism in the context of union elections. Cases were compiled from the use of two legal search engines, Lexis and Westlaw. There are a total of 115 cases decided by the NLRB. (In the following section, I will examine the 43 cases that were decided by federal appellate courts.) As mentioned earlier, the cases are the universe of Board decisions; however, they are not the universe of racist incidents in union election campaigns as I examine only those cases that reach the Board's review and are published.<sup>3</sup> An untold number of cases are either handled only by an administrative law judge or are never officially brought forth by a party as a violation of labor law. But, for the purposes of this paper, these cases allow us to systematically examine the Board's decision making and the principles it uses to resolve cases.

The earliest reported cases (1935–66) took place almost exclusively in the South, and dealt with clear manifestations of what political scientists label “traditional racism” where employers used racially explicit epithets to scare white workers, and on occasion black workers, from voting for the union. With the exception of a few conflicts between unions and Jewish employers and lawyers, all of the cases during this time involve divisions between whites and blacks. The Board handed down decisions in 71 cases during this time period, only two of which responded to an accusation by the employer that the union instigated the racist activity. In some cases, the racism went beyond words to physical violence, leading to the physical assault of black employees and suspected union organizers, the brandishing of weapons, and even the shooting of union members. Shortly after the Civil Rights Act of 1964,

<sup>3</sup> The cases examined are also not the universe of incidents of racism to which the Board has officially responded. The Board has also dealt with union and employer racism in the context of hiring discrimination and a “duty of fair representation.” These matters represent separate realms of cases beyond the scope of this paper and have frequently involved not only federal courts, but other administrative agencies such as the Fair Employment Practices Committee, the Equal Employment Opportunity Commission, and the Department of Labor. (For further discussion of these matters, and the contrasting ways in which these different agencies handled union racism, see Frymer 2004; Kryder 2000.)

the nature of the cases changed dramatically. Out of the 44 cases decided by the Board between 1967 and 2000, only 25% involved accusations that the employer committed the racist act. Instead, the overwhelming number of cases involved situations where the employer accused the union of racism, either through union member epithets against other workers or the employer, or in situations where unions used race-specific appeals to mobilize African American, Asian American, or Latino workers. By the 1980s, an increasing number of these cases involved conflicts between racial and ethnic minorities—blacks and Latinos, Filipinos and Japanese, Jews and Protestants, and so forth. But, despite these differences in the parties involved, differences in which groups were the target of racist actions, and differences in the Board's political composition, the Board has consistently understood these acts as a product of institutional combat.

### Traditional Racism, 1935–66: Employer Threats and Agenda Setting

Two types of scenarios were particularly common during the cases in the first three decades of the Act, almost all of which predate the 1964 Civil Rights Act. More than two-thirds of these cases were situations where the employer attempted to defeat the union by appealing to white worker racism with suggestions that electing a union would lead to a racially integrated workforce. The cases involved a variety of types of ways in which employers race-baited. In one, the employer called the union organizer “a communist, an agitator, and generally a ‘no-good nigger,’” (*California Cotton Oil*, 20 NLRB 540, 549 [1940]); in another, the employer told white workers that if a union came in, the factory would “be fulla Negroes” (*S.K. Wellman Co.*, 53 NLRB 214, 215 [1943]); while a third case involved an employer hiring five African Americans to pass out leaflets supporting the union in an effort to push the company's white workers to vote in the opposite direction because of a perception that the union was associated with racial diversity (*Heintz Division*, 126 NLRB 151 [1960]). The events of *Bush Hog, Inc.* (161 NLRB 1575, 1592 [1961]) typified the period. The company president told workers during the election campaign in Selma, Alabama that “if the Union went in and all that, that we would have to work with Negroes” and if the union were not elected the employer “would keep them out.” The Board described the election campaign as “charged . . . with the atmosphere of the Negro revolution for equality and the march from Selma to Montgomery” and both sides were accused of playing up race issues. The company president was a member of the local White Citizen's Council and a chief part of his campaign against the union was to portray it as pro-civil rights (a charge the Teamsters consistently and vehemently denied to the workers). After referring to the Teamsters' donation of \$25,000 to Martin Luther King, the president “went on to say that we are not going to hire any Negroes. [But] if the Union comes in you will be working beside Negroes.” Later in the

campaign, he called the Teamsters “nothing but nigger-loving gangsters.” A poster was put up by the employer that showed a “fat” African American man smoking a cigar with the caption, “Us and that Union are going to change things around here.”

A second set, representing roughly 20% of the cases during this time, involved situations that paralleled the aforementioned cases but involved employers attempting to scare African American workers, either by intimating that whites would be hired to replace them, by firing African American workers and replacing them with white workers to scare others (*Bess F. Young*, 91 NLRB 1430 [1950]), or through straightforward physical intimidation. Typical was *Fred A. Snow Co.* (41 NLRB 1288 [1942]), where the employer told his black employees that if the union organizing movement were to succeed, he would turn his business over to his son who “didn't like colored people.” Sometimes the employer's actions were fairly subtle, as when an employer brought in roughly 100 white job applicants in full view of the majority black employees just before the employees were to vote on whether to endorse a union as their representative (*Associated Grocers Port Arthur, Inc.*, 134 NLRB 468 [1961]). Others were more explicit. In *Taylor Colquitt Co.* (47 NLRB 225 [1943]) the employer's actions, as well as the actions of white employees opposed to the union, involved direct acts of racism and intimidation. In this case, the employer repeatedly threatened physical harm toward her black workers, telling them early in the campaign that “All you boys will be out of a job; you won't have nothing to do; you will be going around hungry. Furthermore . . . there is going to be some trouble around here if you don't stop this Union. There [will] be some blood shed.” Later, after she confronted a union supporter (he had told her, “Mrs. LaBoone, I don't want to talk to you on this. . . . You are a white lady. I am a colored boy. I couldn't talk to you on nothing like that”), she warned him that “If you vote for it I will kill you.” Later, she encouraged white employees, all brandishing rifles, to confront a group of black workers who were told they would not be allowed to go on “organizin' agin the whites.”

The NLRB members were initially unsure how to respond to these cases and many of the decisions included dissents and concurrences. The NLRA had been passed by Congress with no specific provision that defines racism as illegal or an “unfair labor practice.” A “duty of fair representation,” while endorsed by the Supreme Court in a case involving the parallel National Railway Act (*Steele v. Louisville & Nashville R. Co.*, 323 U.S. 192 [1944]), would not be enforced by the NLRB until 1964 and only involved representation issues for those who were already union members. In some of these early cases, the Board held that the employer's comments, and even acts of physical intimidation and violence, did not influence the election results. The Board argued that as long as the comments were not combined with a clear threat and as long as they did not misrepresent the facts, the statements themselves, while unpleasant and undesirable, were not coercive (e.g., *Happ Brothers Co.*, 90 NLRB 1513 [1950]; *Sharnay Hosiery Mills*, 120 NLRB 750 [1958]).



Particularly in the cases involving white workers who responded to race-baiting, the Board was sympathetic to employer arguments that they were doing nothing more than making factual representations about the union and what would happen if the union were to win. The problem, employers argued, was not their actions but the workers' racist reactions.

But even in these cases, Board members were frequently in disagreement with each other. One concurring member in *Westinghouse Electric Co.* (119 NLRB 117 [1957]) disagreed that employers were just providing factual representations: "The more subtle problem, however, arises when the reference to job retention or job loss is tied to the fact that the Union has a policy, at odds with that of the Employer, which calls for disregarding racial lines in the allocation of jobs, the implication being that, if the Union wins the election, union policy will probably prevail thereafter in the plant." The fear was that employers were playing on worker psychology to either threaten or distract them from whether or not they wanted a union. In other cases, the Board found employer actions to be "calculated to feed upon the employees' latent prejudices and to arouse resentment and antagonism against the Union, [to distort] the Union's policy of equality into a threat" (*Pittsburgh Steamship Co.*, 69 NLRB 1395, 1414 [1946]). In this case, the employer used both physical and verbal threats throughout the campaign to play on the workers' opposition to integration. One manager told the workers, "I'm going to hire a big nigger to be your partner and the blacker the better." A second manager told workers, "The CIO isn't going to last always, President Roosevelt isn't going to live always, and when he dies all the Jews, the God damned Jews are going to be out and we will have a different set-up." Still later, the same official told workers, "if you do win the election, you are going to bring up a lot of goddam niggers from the coast, and they are going to put one in every room. . . . How would you like to eat and sleep with a nigger?" In this case and others like it, the Board found the comments to be threats by the employers that working conditions would worsen if a union came in. Employers were succeeding because their words resonated with the workforce, but the Board focused on the employer's prompting—and recognition that it was designed to defeat the union—and not the individual attitudes.

By the 1960s, Board members were coming to some consensus on how to interpret and respond to racism in union elections and in *Sewell Manufacturing* (138 NLRB 66, 67–68 [1962]) the Board set out a number of broad principles that have since become a standard or "doctrine" by which subsequent union racism would be judged and compared. *Sewell* involved an election in two Georgia towns where the union was soundly defeated because the employer seemingly linked the union organizers to the civil rights movement. Two weeks prior to the election, the employer showed workers a picture of the union president dancing with an African American woman, with a caption referring to it as "race mixing." He pointed out that the union used membership funds to support various civil rights

groups and told the workers "the unions . . . have tried to force (integration) down the throats of the people living in the South." In overturning the election results, the Board made two statements about how the institutional context could impact the manifestation of union racism. First, it emphasized that a union election was different than a regular political election. Not all the participants were equal and the Board believed it had a responsibility to scrutinize speech and behavior by all of those involved to "insure that the voters have the opportunity of exercising a reasoned, untrammelled choice" regarding unionization. Because the employer controls the employment relationship and, in almost all circumstances, possesses more economic power than does the individual employee, the employer's words were to be considered imbued with a "force independent of persuasion" (*NLRB v. Federbush Co., Inc.*, 121 F.2d 954, 957 [2nd Cir., 1941]). Second, the Board made a distinction between when racism was an acceptable and rational part of a campaign and when it was not. Racist language, race-baiting, or other forms of racist speech were potentially allowable and legitimate in union elections. "Some appeal to prejudice of one kind or another is an inevitable part of electoral campaigning." It is only when the racist speech "can have no purpose except to inflame the racial feelings of voters in the election," and particularly when the speech is being imposed in a manner to infringe upon the institutional mandate of the Board—to ensure that elections are independent of coercion—that it would find the speech actionable and overturn the election results. In this case, "it seems obvious from the kind and extent of propaganda material distributed that the Employer calculatedly embarked on a campaign so to inflame racial prejudice of its employees that they would reject the Petitioner out of hand on racial grounds alone."

### Union Racism—Focusing on a Harm Independent of the Act

Although the Board's early decisions dealt almost exclusively with employers' race-baiting in an effort to divide unions, 75% of the cases after 1966 would deal with accusations by the employer that the union had race-baited or that a union member committed a racist act. This immediately suggests an institutional dynamic at work. The Board's decision in *Sewell* was widely discussed in employer manuals; anti-union consultants told employers specifically that they could not race bait. And seemingly, employers stopped race-baiting, at least in the manner of the pre-*Sewell* cases. Interestingly, while the Board found the employer guilty of more than 80% of the cases in the first time period examined, it would only find the union guilty in 4 of the 33 cases that came after 1966. Instead of finding the racist act harmful and actionable as in many of the aforementioned cases, the Board repeatedly dismisses the union acts of racism in the latter cases as either incidental, harmless, or essential and rational to union mobilizing efforts. As the Board stated in *Maple Shade Nursing Home, Inc.* (223 NLRB 1475, 1483 [1976]), a case in which union members frequently

belittled the employer's heavy "Jewish" accent, "union activities (are) not any form of tea party. The Union did nothing here to inflame irrational prejudices, and the employees laughed at their own jokes." In *Bancroft Manufacturing* (210 NLRB 1007 [1974]), it dismissed the relevance of a union comment that a black worker who had been given a car by the employer was a "sold out soul brother," whereas in another case where union supporters repeatedly called a man a "house nigger," the Board found the comments to be "obvious ribbing" and an effort by the union to get the man to "abandon his servant type mentality" (*Vitek Electronics*, 268 NLRB 522 [1984]). When the Board in *Beatrice Grocery Products* (287 NLRB 302 [1987]) looked at a statement by a union representative that a supervisor had called the employees "dumb niggers," it argued that the statement by the union member was made in order to confront racism, not create it.<sup>4</sup> As long as the topic of discussion is "whether employees have been unfairly treated," it is legitimate regardless of the racist content (*Coca-Cola, Inc.*, 273 NLRB 444 [1984]).

By focusing on the impact of racist words, the Board has argued that words themselves are not, by definition, harmful—they can be potentially neutralized by political or institutional context. In *Foundry Div. of Alcon Indus* (328 NLRB 129 [1999]), the Board argued that workers calling each other "nigger" while waiting in line to vote did not have an impact on their behavior because other workers immediately countered the comments. Whereas a federal court (*United Packinghouse v. NLRB*, 416 F.2d 1126 (DC Cir, 1969)) held that racism in union campaigns inevitably led to docility and a demobilization of worker protest, the Board disagreed: racism and discrimination were political categories which could be mobilized and manipulated in a myriad of ways, some of which could be for the good. "A continued practice of discrimination may in fact cause minority groups to coalesce, and it is possible that this could lead to collective action with nonminority group union members" (*Jubilee Manufacturing Co.*, 202 NLRB 272 [1973]). It reiterated this argument in *Handy Andy* (228 NLRB 447 [1977]), claiming that union racism may serve multiple purposes: "employers faced with the prospect of unionization will be provided and have been provided . . . an incentive to inject charges of union discrimination . . . as a delaying tactic in order to avoid collective bargaining altogether rather than to attack racial discrimination." Moreover, the Board consistently looked at whether the statements were made as a part of electoral strategy or in isolation. In *DID Building Services* (291 NLRB 37 [1988]), the Board found that comments made in the heat of the moment were probably "discounted" by workers "as impulsively made." Comments that were "vile and

seething with prejudice" were considered isolated and irrelevant to the election campaign. The point here is not to defend or legitimate racist acts and practices, nor to disagree with those that argue words alone can "wound" (Feagin 1991; Mackinnon 1993; Matsuda 1993). It is to argue that such an act is fundamentally situated within a broader set of politics and can only be understood within this context before it is deemed actionable.

When the Board has found union racism to be worthy of overturning an election victory, it has involved situations where the racism replaced the political confrontation as the primary focus of debate and where union leaders clearly acted strategically in placing the race issue on the agenda. Two contrasting cases provide an example. For instance, in *YKK (USA)* (269 NLRB 82 [1984]), the Board confronted a situation where the comments were clearly made by union leaders and became a centerpiece of the union's campaign. Union leaders at a Japanese-owned zipper company passed out campaign literature that made repeated derogatory references aimed at the owner's nationality. At a union meeting shortly before the vote, the union's national representative told the workers to stick together against the "Japs," ending his speech with words to the effect that "we beat the Japs after Pearl Harbor and we can beat them again." Later, this same union officer shouted at a Japanese engineer of the company, "[t]here goes one of those damn Japs. Go back where you came from, you damn Jap," while the union vice president wore a t-shirt with the phrases "Japs go home," and "slant eyes." The Board held that the racism in this case was distinguishable from past cases because "[t]here is no conceivable way that a reference to beating 'Japs' at Pearl Harbor could be relevant to a legitimate campaign issue." Because the union made race the center of the campaign, the Board viewed the racism as unconnected to legitimate worker concerns and held that it served no purpose beyond being inflammatory and illicit and an effort to mobilize workers around their racism towards the Japanese. In contrast, the Board allowed an election to stand in *KI (USA)* (309 NLRB 1063 [1992]), a case where union members again attacked the Japanese company owners, both with private jokes among employees and by disseminating and attacking a letter that it claimed to be from the company's president as an example of the Japanese "screwing us over." The disseminated letter stated: "I am appalled at the typical lazy, uneducated American worker. . . . I suggest the Americans start developing a healthy respect for Japan because one of my colleagues will eventually become your boss." The Board distinguished this case from *YKK*, arguing that "notwithstanding any racial overtones, the topic of how American workers were regarded by management was a relevant campaign issue."

<sup>4</sup> A dissenting Board member wrote in response, "The remark was such that the employees were not likely soon to forget it . . . The history of the term 'nigger' has rendered the use of it so opprobrious that it triggers instantaneously a whole complex of memories and resentments. We may as well ignore the devastating effects of a discharged firearm by describing the pull of the trigger as 'isolated' as pass silently by the effects the use of this single word is capable of causing."

[T]he context was that the employees were concerned about the impact of the attitudes of the Japanese owners on their workplace. Thus, (dissemination of the letter) appears to be an attempt at least to pose the question of whether there is some connection between the two. It

does not automatically follow that this communication is inherently objectionable. Although such claims raise the specter that some voters may overreact and respond in an equally prejudicial manner . . . the Board has not equated the broaching of such topics to opening a Pandora's box.

### Contrasting Racism as Mobilizing versus Racism as Dividing

The Board has responded in an entirely different manner to cases where employers have accused minority workers of using race as a mobilizing tool in their election campaigns. In *Aristocrat Linen Supply Co.* (150 NLRB 1448 [1965]), African American workers used civil rights appeals to promote worker solidarity against the employer. The union passed out a flier to a predominantly black workforce that ended with the statement "this is why the labor hater is always a twin-headed creature spewing anti-Negro talk from one mouth and anti-propaganda from the other." Union leaders later exhorted workers not to be a "Handkerchief head Uncle Tom." The Board, while finding the campaign rhetoric "undeniably based upon a racial issue," argued that "a distinction must be drawn between racial propaganda designed to inflame racial hatred and set the tone of a union campaign as a battle of one race against another as in *Sewell*, and racial propaganda designed to encourage racial pride and concerted action." The same year, in a case of similar circumstances, the Board again contextualized race and racism within the political battle, allowing for multiple ways in which race-specific campaigns can be used: "An appeal to racial self-consciousness may produce a variety of emotions, depending upon the context. In some cases, such appeals may result in vicious race hatred. In another circumstance, such appeals may promote reasoned and admirable ambition in an unfortunate race of people" (*Archer Laundry*, 150 NLRB 1427 [1965]). In a later case where the union mobilized around race issues, *Baltimore Luggage* (162 NLRB 1230, 1233 [1967]), union organizers told black workers that they received lower pay than white workers because of their race. Again, the Board distinguished between the irrational use of race language in *Sewell* and the arguably rational way in which it was presented here: "In *Sewell*, we did not lay down the rule that parties would be forbidden to discuss race in representation elections." The Board argued that unions could make race-specific appeals when they are used to promote the rights of disadvantaged groups in their quest for economic empowerment: "campaign material of this type is directed at undoing disadvantages historically imposed [generally unlawfully] upon Negroes because of their race, through an appeal to collective action of the disadvantaged. The choice of racial basis for concerted action has been made, not by the victims who organize to seek redress, but by those who use race as a basis to impose the disadvantage."

In *Carrington South Health/Care Center* (1994 NLRB Lexis 397 [1994]), a largely African American workforce was given three cartoons by union leaders designed to encourage their support for the union. Two

cartoons showed a clearly white owner either exploiting or enslaving a clearly black employee, while the third showed a white "boss" directing a nervous looking black employee to an electric chair, stating "You don't need your union rep. Just have a seat and we'll discuss your grievance like two rational human beings." The Board found these race-specific cartoons to be appropriate for an election because they reflected the benefits of being in a union—in fact, the Board argued that these were not race-specific appeals at all, but simply a form of contestation over economic concerns. In *Bancroft Manufacturing Co.* (210 NLRB 1007, 1008 [1974]) the Board dealt with a case in which a black union organizer told workers to stay in solidarity because individual black workers were being bribed with new cars, and because if the union lost, "all blacks would be fired." Here the Board doubted whether the union would make strategic use of such racially specific comments because the workforce was nearly 60% white. Since it would be "suicidal" to play the race card in this way, and since the use of race during the campaign stressed "black pride, the past history of discrimination against blacks in American society or the present disadvantaged status of blacks as a class," it found the comments to be a legitimate part of the campaign discourse. In 1998, as race mobilization cases became more and more frequent, and as employers continually objected to their use by referring to the *Sewell* doctrine, the Board's General Counsel, William Gould, proposed a new doctrine that would be used to distinguish the racial mobilization cases from other racist acts in union campaigning:

Because the employer controls the employment relationship and . . . possesses more economic power than does the individual employee, the Board's concerns about racial appeals expressed in *Sewell* . . . have peculiar applicability to remarks of employers as opposed to those of unions and their representatives. . . . Union organizational efforts aimed at blacks and other racial minorities and women must necessarily focus, in part, upon grievances peculiar and unique to such groups, i.e., employment conditions which are attributable to racial inequities or what appear to be racial inequities and other forms of arbitrary treatment (*Shepherd Tissue*, 326 NLRB 369 [1998]).

As we will see, although this may seem an unsurprising interpretation by the Board given the comparison with the cases that it has found objectionable, the race mobilization cases will provide one of the most dramatic discrepancies between the Board's understanding of race and the understanding of federal courts.

### Collective Action

We saw previously how the Board scrutinized the institutional dynamics that led both sides to attempt to shift the focus of a union campaign from politics and economics to race. Another way in which the Board scrutinizes the institutional context of racist acts involves its examination of collective action concerns within a union organizing drive. Collective action problems in unions are one of the most fundamental ways in which

labor law is different from the types of cases that ordinarily appear before federal courts. As mentioned earlier, unlike most other realms of law, the individual is not at the center of labor law. Unions exist because workers agree to limit their individual opportunities in the effort to benefit as a group. As a result, unions consistently confront the difficulty of maintaining the support of potential “free riders” who may choose to reap the benefits of the union without participating in the costs of its formation and maintenance (Olson 1971). The NLRA is cognizant of this and there are numerous statutory ways that unions can discipline individuals. As the Supreme Court wrote in *NLRB v. Allis-Chambers Mfg.* (388 U.S. 177 [1967]):

National labor policy has been built on the premise that by pooling their economic strength and acting through a labor organization freely chosen by the majority, the employees . . . have the most effective means of bargaining for improvements in wages, hours, and working conditions. The policy therefore extinguishes the individual employee's power to order his own relations with his employer and creates a power vested in the chosen representative to act in the interests of all employees.

The nature of collective action, and the fact that unions are designed to promote a “public good” to overcome economic inequalities, creates further problems for the organization in maintaining internal hierarchies and leadership (Levi 2003; Mansbridge 1986; Polletta 2002). Unlike a company that is run by a “boss” or “CEO,” union hierarchies are relatively fluid and democratic. The power of union leaders to keep members disciplined is less fundamental than that of a CEO, and as a result members are not controlled by leaders when they wish to “speak” on behalf of the union. Particularly in union organizing battles where many of the union supporters are not official union members until after a certifying election, unions face significant problems in maintaining coordination and a unified message. As a result, labor law has provided unions with different opportunities to promote a group identity in spite of collective action problems. Labor law allows unions (at times) to impose closed shops, to punish members (within limits) for refusing to follow majoritarian decisions, and generally to prevent its members from dissenting and abstracting themselves from the union decision-making process. Today, although some of these opportunities have been weakened or even taken away, the NLRB continues to recognize the collective action problems that underlie union leadership and action.

More specific to questions of handling and understanding racism, the Board has argued that it will only find a labor union accountable for a racist act of a union supporter or member if the actor is deemed an official leader and authorized to comment on the union's behalf. Other individuals, even strong union supporters, are generally deemed beyond the union's control and responsibility. As a result, the Board applies a less rigorous “third party” standard where the union leadership cannot be connected to the racial statements made by non-union leaders—even if they are employees who

are close to the organizing campaign. Similar to the other examples mentioned earlier, then, the Board's separation between leadership and members reflects that not all racist acts are equal as well as recognition that institutional relationships will otherwise make it more likely that a union member is involved in a racist act than a member of the employer's staff. That the cases post-1966 have been so heavily dominated by accusations against the union is arguably reflective of this dynamic. Since the late 1960s, employers have increasingly relied on “union-busting” consultants that specify what employers and their managers can and cannot say and do during a union campaign. Employers simply have more hierarchy and discipline over their managers and supervisors—they can fire managers, for instance, and suppress dissent—than unions who by law must protect employee speech and dissent, are limited in the forms in which they can discipline their members, and powerless against union “supporters” who are not currently members.

The cases reflect recognition of this political and institutional inequality. For instance, in *Zartic, Inc.* (315 NLRB 495, 500–508 [1994]), the Board found that statements by a union member linking the employer to the Ku Klux Klan (KKK), while baseless and designed intentionally to “exploit the ethnic fears of the Hispanic employees by making a visceral connection between the KKK and working conditions,” were nonetheless not liable under the *Sewell* doctrine because the union lacked control over the individual. The employer had a “stricter burden of proof . . . to establish that the conduct of third parties was of so serious a nature.” The Board similarly discounted comments by African American organizers (“Boy, you white sons-of-bitches, you are all the same, you're scared to take a stand”) as being outside of the authority of the union leadership (*Herbert Halperin Distributing Co.*, 1968 NLRB 247 [1986]) and, in *Air Express Int'l Corp.* (289 NLRB 608 [1988]), argued that when a pro-union employee told others of the employer's dislike of Cubans, that the comment was not a “systematic attempt (on the part of the union leadership) to inject the ‘racial’ issue into the campaign; but that the employees probably had blown the statement out of proportion in the retelling of it.” Many other cases, meanwhile, dealt with rumors that workers spread during the course of the election campaign. In one, a disagreement between a labor board member and the employer's attorney was misrepresented by workers as a disagreement over whether workers would be allowed to speak Spanish at work, and thus on the eve of the election, workers discussed widely the belief that the employer was anti-Latino (*Singer Co.*, 191 NLRB 179 [1971]). Though it injected racial animus into the campaign, it was deemed outside the responsibility of any union actor. More typical of the rumor cases was *Information Magnetics* (227 NLRB 1493 [1977]), where workers spread rumors that the employer had brought in the Immigration and Naturalization Services to deport union supporters who were illegal immigrants. Again, although the rumor clearly had an impact on the election, the Board held that the rumors were not controllable by union

leaders. In a contrasting case, when a union leader threatened a worker with deportation, the election was overturned (*Professional Research, Inc.*, 218 NLRB 96 [1975]).

But even when union leaders are involved in the racist action, the Board has scrutinized their participation and the degree to which they officially spoke for the union and which other union leaders had control over the individual. In *Benjamin Coal* (294 NLRB 572 [1989]), where the Board dealt with comments made by members of the union organizing committee, it argued that because the committee was made up of volunteers and was open to any employee who wished to join, the union leadership could be distinguished from its organizers. It held that the union did not “echo or condone these highly offensive sentiments.” When union leaders heard antisemitic statements during the campaign coming from a union organizer, “the organizers immediately quieted (him) and told the audience that such comments were irrelevant to the campaign.” Although the Board made clear that racist statements had no place in a campaign, “To hold that the election was tainted by such prejudice would be to hold that no election could ever be held in any plant with a prejudiced work force unless the union attempting the campaign were able to accomplish what management itself had been unable to do before the union came on the scene, namely, eliminate all expressions of racial, ethnic, or religious bias.” In *Pacific Micronesia Corp.* (326 NLRB 458 [1998]), meanwhile, the Board, in refusing to overturn the election results, pointed to the fact that the president of the union refuted statements made by a union organizer to the effect that the company was hiring Nepalese workers to weaken the strength of the company’s Filipino workers.

### Federal Courts: the Inherent Damage and Irrationality of the Racist Act

After the Board makes the initial decision on election conduct, either side can appeal to a federal appellate court. Sometimes there is no appeal and sometimes the appeal centers on a different accusation of unfair labor practices—in many of the cases discussed here, racism is just one of many changes that the Board and courts are asked to deal with. When a Board ruling on race is appealed, more often than not, federal courts defer to Board decisions, reflecting the deference they generally give to administrative agencies. Nonetheless, federal courts have differed in significant ways with the Board’s interpretation of racism in union elections, and the manner of this clash is theoretically illuminating, as it has reflected a very different understanding of racism, one that parallels the individual-prejudice model so widely endorsed in political science. When federal judges object to NLRB decisions, it is consistently on the same grounds; that racism is itself the harmful act and that it is irrelevant whether the appeal to racial passions was made by the employer with the goal of dividing the workforce, or by the union with the goal of enhancing solidarity, or whether it was

made by a worker who was not a member of the union leadership. To these judges, racist acts are never to be tolerated and are always the responsibility of the individual who carried them out. Federal judges have argued that racism is intrinsically irrational and thus can never be understood differently regardless of the context and institutional dynamic in which the racist act was situated. Prejudice trumps institutional considerations.

An emblematic example of federal court interpretation of union racism is when the Sixth Circuit (76 F.3d 802, 807 [1996]) reviewed and overturned the *Carrington* case discussed earlier. The Board had found the acts to be a way of mobilizing disadvantaged workers around issues that intersected race, class, and power, and argued that they were “devoid . . . of appeals to racial bigotry.” The federal court disagreed, holding that the racially specific cartoons that identified African American workers being exploited by their employer were used deliberately by the union to exacerbate racial feelings with irrelevant and inflammatory appeals. Although two of the cartoons made a passing reference to legitimate campaign issues, the judge found the imagery to be “quite troubling” and a “graphic appeal to racial prejudice.”

Each cartoon uses obvious images of bondage or violence visited upon racial minorities by a white majority: a white man purchases a group of black (or mostly black) workers; a group of workers labor as beasts of burden, pulling their superiors in a wagon while being whipped; a black worker is to be summarily executed by a white overlord. . . the cartoons could therefore be construed as a deliberate exacerbation of racial feelings by irrelevant and inflammatory appeals.

Other federal courts have reacted in like fashion, responding to the inherent irrationality of the act and not the political context in which the manifestation occurred. The Sixth Circuit overturned *KI Corp.*, as discussed earlier (where the union pointed out a racist letter written by a Japanese business owner intended toward the employees), because the “negative stereotyping . . . has [no] legitimate place” regardless of context and the use of the letter “exceeds the bounds of legitimate discussion” (*KI Corp. v. NLRB*, 35 F.3d 256 [1994]). The Fourth Circuit in *NLRB v. Schapiro & Whitehouse, Inc.* (356 F.2d 675 [1966]) found that a union leaflet that pointed out that a union would help solve racial discrimination in employment was “deplorable” and “highly inflammatory” speech. The court wrote that the “equality of race [was] not presently an issue. That the majority of the employees were Negroes did not make it so.”

Union antisemitism has been the issue of a series of cases where federal courts have overturned Board decisions. In one, where the employer was compared to Hitler, the Fourth Circuit overturned the election on grounds that the union had “interjected into the election one of the most sordid episodes in modern history” (*Schneider Mills, Inc. v. NLRB*, 390 F.2d 375 [4th Cir., 1968]). The Third Circuit (*NLRB v. Silverman’s Men’s*

*Wear, Inc.*, 656 F.2d 53 [3rd Cir., 1981]) overturned the Board in a case involving anti-Jewish rhetoric when a union representative called the company's vice president a "stingy Jew." The court held that appeals to racial prejudice constituted a prima facie warrant for setting aside an election. Unlike the Board, then, the explicit assumption of the court was that racism had no economic or political component that in any way could be associated with an organizing drive. In this case, the court found "such a remark has no purpose except blatantly to exploit religious prejudices of the voters. It deliberately injects 'an element which is destructive of the very purpose of an election.' . . . We can see no reason for the remark except to inflame and incite religious or racial tensions." Similarly, in *Katz v. NLRB* (701 F.2d. 703 [7th Cir., 1983]), the court overturned the Board for not finding antisemitic statements to be problematic: "There is no conceivable way in which either the movie 'Holocaust' or the Nazis' treatment of Jewish people during World War II could be relevant to a legitimate campaign issue. To the extent that the priest's comment regarding the wealth of Jewish people, as juxtaposed to the poverty of the employees, might be relevant to the campaign, the point could have been made without resort to a religious slur."

Federal courts have also been far more likely to discount consideration of collective action issues as has been done by the Board. In the aforementioned *Katz* decision, the court dismissed the Board's determination that the statement had been made by a priest with no direct affiliation with the union. The court held that the collective action questions raised by the Board were not determinative; all that mattered was whether the statements were racist and provocative and hence overturned the union victory. The Eleventh Circuit (*M&M Supermarkets v. NLRB*, 818 F.2d 1567 [1987]) made no distinction between the union and its leadership and an "outspoken union supporter and advocate" who said in support of the union: "the damn Jews who run this Company are all alike. They pay us pennies out here in the warehouse, and take all their money to the bank. . . . Us blacks were out in the cotton field while . . . the damned Jews, took their money from the poor hardworking people." The Board had concluded in contrast that he was not a union agent, nor that there was any evidence that the union authorized or was even aware of his actions. The court however stated simply that "such feelings [have] no place in our system of justice." Similarly, in *NLRB v. Eurodrive, Inc.* (724 F.2d. 556 [6th Cir., 1984]), the court overturned a Board's decision that had found a union not to be responsible for an organizer's statements, nor responsible for what it perceived to be racial tensions that existed in the workplace before the beginning of union organizing. While the Board found the comments made by the employee to be "intemperate, abusive, and inaccurate," the court argued that the organizer's attenuated link to the union was less relevant than the racial statement itself, and that the statement had exacerbated racist tensions and "had an appreciable effect upon the employees' freedom of choice."

## CONCLUSION

I have argued that the Board's approach to racism in union elections provides theoretical insights into a more complicated and complete understanding of the dynamics that lead individuals to commit racist acts. The NLRB's approach, by scrutinizing context, differentials in power, and the particular use of agenda setting and collective action issues, places the individual act within a broader political and institutional setting. It highlights how institutions impact actors, biasing the ways in which racism is manifested and the ways in which it is hidden. To have a fuller understanding of racism, then, we need to place it in a broader political-institutional context in which individuals act strategically to pursue goals.

I do not wish to argue that the Board's approach is the only explanation of racism, nor that its holdings in each case are always "correctly decided." The Board's approach is both novel and an important addendum to those who too often limit their understandings to individual psychology and attitudes. But the Board's particular application is not the only way to interpret racism within a political and institutional context. The Board members themselves are institutionally limited in the ways that they can interpret and address discrimination in unions. The development of labor institutions in the United States occurred at a time when the dominant national unions of the American Federation of Labor were overwhelmingly white and male. Key union leaders pushed successfully for provisions of the NLRA that enabled them to maintain racial hierarchies and these provisions were supported by southern Democrats in Congress (Frymer 2003). It took three decades after its creation, moreover, for the NLRB to sanction a union on grounds that it was discriminating in the hiring and representation of minority workers, in part because the Board believed it had previously lacked the statutory power to do so. The NLRB's "tardy assumption of jurisdiction" over fair representation issues involving race led the Supreme Court to refuse to allow the Board to maintain sole control over labor civil rights matters, particularly involving racial discrimination under Title VII (*Vaca v. Sipes*, 386 U.S. 171, 183 [1967]). The Board has also been largely absent from handling widespread discrimination on the basis of nonracial characteristics such as gender and disability, again in part because it does not see itself as an institutional actor designed to handle such matters (indeed, with the exception of a broad duty of "fair representation" to all workers, there are no specific provisions in labor law on topic).

Perhaps symbolic of the Board's institutional limits in handling race-specific discrimination is a case where it defended the principle of a union's right to have exclusive representation in the workplace. In doing so, it allowed the firing of a group of African American workers who protested their company's discriminatory employment practices and what they perceived to be the unwillingness by their union to adequately represent them (*The Emporium and Western Addition*



*Community Organization*, 192 NLRB 173 [1971]). The Board thus confronted a conflict between principles of civil rights and union authority and since the NLRA emphasizes the latter over the former, it is perhaps not surprising that the Board found on behalf of the union. The Board held that the union was the only representative that could bargain with the employer under labor law; because the workers protested the company's civil rights policy independent of the union's support, they were legitimately fired from their jobs for participating in an illegal work stoppage. The Board was initially overturned by the D.C. Court of Appeals on grounds that labor laws should not intervene when principles of racial equality are at work. Civil rights, the federal court argued, trumps considerations of union power. But Thurgood Marshall, writing for the majority of the Supreme Court (*Emporium Capwell v. Western Addition Community Organization*, 420 U.S. 50, 67 [1975]), sided with the Board. Whereas the appeals court had argued that confronting racism should not be obstructed by labor law statutes, Marshall countered that questions of race and labor power were inseparable: "Competing claims on the employer's ability to accommodate each group's demands . . . could only set one group against the other. Having divided themselves, the minority employees will not be in position to advance their cause unless it be by recourse seriatim to economic coercion, which can only have the effect of further dividing them along racial or other lines." Limiting union rights, Marshall argued, would undermine the ability of unions and their workers to fight for civil rights and thus hurt their efforts at ending racial discrimination in the workplace.

*Emporium Capwell* was in many ways emblematic of the debate between the individual and institutional approaches, as well as suggestive of the problems of both the federal court and the NLRB approach to understanding racism. As developed, labor law and civil rights law in America have suggested either/or alternatives, and both have had great difficulty in incorporating racial and class inequality into one regulatory body and one legal understanding (Forbath 1999; Frymer 2004; Iglesias 1993; Katznelson 1989; Klare 1982). The election cases discussed earlier provide some suggestive ways in which this intersection might be accomplished—but this is only meant as a starting point, and future work in applying a political approach to these questions must go further in intersecting questions not just of race and class, but of gender, sexuality, and other existing dynamics of inequality (for very good starts at this, see Cohen 1999; Warren 2004; Young 1990). In particular, it is at the intersection of issues involving the contestation between marginal groups that the individual model of prejudice becomes most wanting. Although historically varied, institutions have quite often created opportunities for powerful actors to benefit from those less powerful being pitted against each other in sites where these intersectional conflicts are most visible. It is when conflict between less powerful groups is most intense that an institutional explanation can allow us to step back and see

such motivations and behavior in a broader context of power, rationality, and structure (Piven and Cloward 1978).

To see racism institutionally, in turn, suggests a further research agenda for political scientists beyond the sphere of labor union dynamics. Political-institutional arguments have been made extensively in non-race-specific spheres of politics, where rational choice and new institutional scholars have examined the motivations behind the decision making of individual actors. The assumption in these studies is that psychology is more or less irrelevant to understanding individual behavior. All actors are assumed to be rational, informed, and to act according to the incentives that institutions provide. But in the realm of behavior deemed "irrational"—violence, prejudice, collective action—scholars far too often attempt to explain the phenomenon solely in psychological ways, ignoring how individuals and their leaders are often motivated by the same political and institutional understandings that motivate members of Congress, executives, and interest groups (for notable exceptions to this, see Chong 1991; Robin 2004). The consequence is that racism becomes understood as an innate 'evil' that works in the underbelly of society—it is removed from politics and we thus lose focus of the myriad ways in which strategic politicians and institutions can both promote and prevent racist activity.

A limited number of studies have begun to examine how institutional dynamics influence elite handling of race and racism, focusing on the incentives followed by elected officials that lead them to employ specific types of race strategies (e.g., Fraga and Leal 2004; Frymer 1999; Kim 2001; Walters 1988). This is particularly the case where race intersects with congressional representation, in which scholars interested in strategic behavior view the manipulations and treatment of race as part of normal politics (e.g., Canon 1999; Tate 2004). Among racial attitude scholars, there has been an increasing effort to situate racist acts within institutions. Both James Glaser (2002) and Tali Mendelberg (2001), for example, argue that white attitudes are often ambivalent and individuals are heavily influenced by ballot structures, party leader decisions, and other institutional dynamics. Yet other scholars have attempted to look at broader socioeconomic context to understand the moments of violence and racial protests of the post-Civil Rights era (e.g., Olzak 1992; Sears 1994). And one need not exclude psychological dynamics to incorporate an institutional-political understanding. Psychologists have provided nuanced evidence that shows how almost anyone can be induced to follow orders or change their behavior given a specific institutional context, whether through simple peer pressure or an effort to follow structured rules (e.g., Asch 1958; Milgram 1974; Zimbardo 1973). It is hoped then that by recognizing that racism, like other behaviors in society, can be analyzed as a political act, we will begin to provide a more complete account of why it remains a far too meaningful and widely used form of combat.

## REFERENCES

- Appiah, Kwame Anthony. 1990 "Racisms" In *The Anatomy of Racism*, ed. David Theo Goldberg. Minneapolis: Minnesota University Press.
- Arnesen, Eric. 2001. *Brotherhoods of Color Black Railroad Workers and the Struggle for Equality*. Cambridge: Harvard University Press.
- Asch, Solomon E. 1958 "The Effects of Group Pressure upon the Modification and Distortion of Judgments." In *Readings in Social Psychology*, ed. Eleanor E. Maccoby, Theodore E. Newcombe, and Eugene L. Hartley. New York: Holt.
- Balibar, Etienne. 1992. "Racism and Nationalism." In *Race, Nation, Class Ambiguous Identities*, ed. Etienne Balibar and Immanuel Wallerstein. New York: Verso.
- Balkin, Jack M., and Reva B. Siegel. 2003. "The American Civil Rights Tradition: Anticlassification or Antisubordination." *University of Miami Law Review* 58 (October). 9-33.
- Becker, Craig. 1993. "Democracy in the Workplace: Union Representation Elections and Federal Labor Law." *Minnesota Law Review* 77 (February): 495-587.
- Bobo, Lawrence D. 1988 "Group Conflict, Prejudice, and the Paradox of Contemporary Racial Attitudes" In *Eliminating Racism Profiles in Controversy*, ed P. A. Katz and D. A. Taylor. New York: Plenum.
- Bonilla-Silva, Eduardo. 1996 "Rethinking Racism. Toward a Structural Interpretation." *American Sociological Review* 62 (June): 465-80.
- Brest, Paul. 1976 "In Defense of the Anti-Discrimination Principle." *Harvard Law Review* 90. 1.
- Brown, Michael K., Martin Carnoy, Elliott Currie, Troy Duster, and David P. Oppenheimer. 2003. *Whitewashing Race. The Myth of a Color-blind Society*. Berkeley: University of California Press.
- Canon, David T. 1999. *Race, Redistricting, and Representation The Unintended Consequences of Black Majority Districts*. Chicago: University of Chicago Press.
- Carmichael, Stokely, and Charles Hamilton. 1967 *Black Power: The Politics of Liberation in America*. New York: Vintage.
- Chong, Dennis. 1991 *Collective Action and the Civil Rights Movement*. Chicago: University of Chicago Press.
- Cohen, Cathy J. 1999 *The Boundaries of Blackness: AIDS and the Breakdown of Black Politics*. Chicago: University of Chicago Press.
- Cox, Oliver Cromwell. 1948. *Caste, Class, and Race*. New York: Monthly Review Press.
- de Tocqueville, Alexis. 2001 [1835] *Democracy in America*. New York: Signet Press.
- Draper, Alan. 1994. *Conflict of Interests: Organized Labor and the Civil Rights Movement in the South, 1954-1968*. Ithaca: ILR Press.
- Du Bois, W. E. B. 1935 *Black Reconstruction*. New York: Harcourt Brace.
- Du Bois, W. E. B. 1999 [1903] *The Souls of Black Folk*, ed. H. L. Gates, Jr. and T. H. Oliver. New York: W. W. Norton.
- Feagin, Joe R. 1991. "The Continuing Significance of Race: Antiracist Discrimination in Public Places" *American Sociological Review* 56(1): 101-16.
- Fields, Barbara J. 1982. "Ideology and Race in American History" In *Region, Race, and Reconstruction: Essays in Honor of C. Vann Woodward*, ed. J. M. Kousser and J. M. McPherson. New York: Oxford University Press.
- Forbath, William E. 1999 "Caste, Class, and Equal Citizenship." *Michigan Law Review* 98(1): 1-91.
- Fraga, Luis Ricardo and David L. Leal. 2004 "Playing the 'Latino Card': Race, Ethnicity, and National Party Politics," *Du Bois Review* 1(2): 297-317.
- Freeman, Alan. 1978 "Legitimizing Racial Discrimination through Anti-Discrimination Law" *University of Minnesota Law Review* 62: 1079.
- Frymer, Paul. 1999. *Uneasy Alliances: Race and Party Competition in America*. Princeton: Princeton University Press.
- Frymer, Paul. 2003. "Acting When Elected Officials Won't: Federal Courts and Civil Rights Enforcement in U.S. Labor Unions, 1935-85." *American Political Science Review* 97(3): 483-99.
- Frymer, Paul. 2004. "Race, Labor and the Twentieth-Century American State." *Politics and Society* 32(4): 475-509.
- Gaventa, John. 1980 *Power and Powerlessness. Quiescence and Rebellion in an Appalachian Valley*. Urbana: University of Illinois Press.
- Getman, Julius G., Stephen B. Goldberg, and Jeanne B. Herman. 1976. *Union Representation Elections. Law and Reality*. New York: Russell Sage Foundation.
- Gilroy, Paul. 1984. "There Ain't No Black in the Union Jack." *The Cultural Politics of Race and Nation*. Chicago: University of Chicago Press.
- Glaser, James M. 2002. "White Voters, Black Schools: Structuring Racial Choices with a Checklist Ballot." *American Journal of Political Science* 46(1): 35-46.
- Goldfield, Michael. 1997. *The Color of Politics Race and the Main-springs of American Politics*. New York: New Press.
- Gould, William B. 1977. *Black Workers in White Unions. Job Discrimination in the United States*. Ithaca, NY: Cornell University Press.
- Gourevitch, Philip. 1999 *We Wish to Inform You that Tomorrow We Will be Killed with Our Families: Stories from Rwanda*. New York: Picador.
- Green, Donald P., Dara Z. Strolavitch, and Janelle S. Wong. 1998. "Defended Neighborhoods, Integration, and Racially Motivated Crime." *American Journal of Sociology* 104(2): 372-403.
- Hall, Stuart. 2000 "Old and New Identities, Old and New Ethnicities." In *Theories of Race and Racism*, ed. Les Black and John Solomos. New York: Routledge.
- Haney López, Ian F. 2000 "Institutional Racism. Judicial Conduct and a New Theory of Race Discrimination" *Yale Law Journal* 109 (June): 1717-1884.
- Hill, Herbert. 1985. *Black Labor and the American Legal System: Race, Work, and the Law*. Madison: University of Wisconsin Press.
- Hochschild, Jennifer L. 1984 *The New American Dilemma: Liberal Democracy and School Desegregation*. New Haven: Yale University Press.
- Iglesias, Elizabeth M. 1993. "Structures of Subordination: Women of Color at the Intersection of Title VII and the NLRA. Not!" *Harvard Civil Rights-Civil Liberties Law Review* 28: 395.
- Katznelson, Ira. 1973. *Black Men, White Cities: Race, Politics, and Migration in the United States, 1900-1930 and Britain, 1948-1968*. New York: Oxford University Press.
- Katznelson, Ira. 1989. "Was the Great Society a Lost Opportunity?" In *The Rise and Fall of the New Deal Order, 1930-1980*, ed. Steve Fraser and Gary Gerstle. Princeton: Princeton University Press.
- Kelley, Robin D. G. 1990. *Hammer and Hoe: Alabama Communists During the Great Depression*. Chapel Hill: University of North Carolina Press.
- Kersch, Ken I. 2004. *Constructing Civil Liberties Discontinuities in the Development of American Constitutional Law*. New York: Cambridge University Press.
- Kim, Claire Jean. 1999 "The Racial Triangulation of Asian Americans" *Politics and Society* 27(1): 105-38.
- Kim, Claire Jean. 2000. *Butter Fruit. The Politics of Black-Korean Conflict in New York City*. New Haven: Yale University Press.
- Kim, Claire Jean. 2001 "Managing the Racial Breach. Clinton, Black-White Polarization, and the Race Initiative" *Political Science Quarterly* 117(1): 55.
- Kinder, Donald R., and Lynn M. Sanders. 1996. *Divided by Color. Racial Politics and Democratic Ideals*. Chicago: Chicago University Press.
- King, Desmond. 1997 *Separate and Unequal: Black Americans and the U.S. Federal Government*. New York: Oxford University Press.
- King, Desmond S., and Rogers M. Smith. 2005. "Racial Orders in American Political Development." *American Political Science Review* 99(1): 75-92.
- Klare, Karl E. 1982. "The Quest for Industrial Democracy and the Struggle against Racism: Perspectives from Labor Law and Civil Rights Law" *Oregon Law Review* 61: 157.
- Knowles, Louis L., and Kenneth Prewitt. 1969 *Institutional Racism in America*. Patterson, NJ: Prentice Hall.
- Krieger, Linda Hamilton. 1995. "The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity." *Stanford Law Review* 47 (July): 1161-1248.
- Kryder, Daniel. 2000 *Divided Arsenal Race and the American State During World War II*. New York: Cambridge University Press.

- Lawrence, Charles R., III. 1987. "The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism." *Stanford Law Review* 39 (January): 317-88.
- Levi, Margaret. 2003. "Inducing Preferences within Organizations. The Case of Unions." University of Washington.
- Lieberman, Robert C. 1998. *Shifting the Color Line: Race and the American Welfare State*. Cambridge: Harvard University Press.
- Lowi, Theodore J. 1969. *The End of Liberalism: Ideology, Policy, and the Crisis of Public Authority*. New York: W. W. Norton.
- Mackinnon, Catharine A. 1993. *Only Words*. Cambridge: Harvard University Press.
- Mansbridge, Jane J. 1986. *Why We Lost the ERA*. Chicago: University of Chicago Press.
- March, James G., and Johan P. Olsen. 1984. "The New Institutionalism: Organizational Factors in Political Life." *American Political Science Review* 78(3): 734-49.
- Marx, Anthony W. 1998. *Making Race and Nation: A Comparison of South Africa, the United States, and Brazil*. New York: Cambridge University Press.
- Matsuda, Mari J. 1993. "Public Response to Racist Speech: Considering the Victim's Story." In *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*, ed. Matsuda, Mari J., Charles R. Lawrence, III, Richard Delgado, and Kimberlè Williams Crenshaw. Boulder: Westview Press.
- Memmi, Albert. 1999. *Racism*. Minneapolis: University of Minnesota Press.
- Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages and the Norm of Equality*. Princeton: Princeton University Press.
- Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. New York: Perennial classics.
- Myrdal, Gunnar. 1944. *An American Dilemma: The Negro Problem and Modern Democracy*. New York: Harper and Row.
- Nelson, Bruce. 2001. *Divided We Stand: American Workers and the Struggle for Black Equality*. Princeton: Princeton University Press.
- Olson, Mancur. 1971. *The Logic of Collective Action*. Cambridge: Cambridge University Press.
- Olzak, Susan. 1992. *The Dynamics of Ethnic Competition and Conflict*. Stanford: Stanford University Press.
- Omi, Michael, and Howard Winant. 1994. *Racial Formation in the United States*. New York: Routledge.
- Piven, Frances Fox, and Richard A. Cloward. 1978. *Poor Peoples Movements: Why They Succeed, How They Fail*. New York: Vintage.
- Polletta, Francesca. 2002. *Freedom is an Endless Meeting: Democracy in American Social Movements*. Chicago: University of Chicago Press.
- Post, Robert. 2000. "Prejudicial Appearances. The Logic of Antidiscrimination Law." *California Law Review* 89: 1-40.
- Reed, Adolph, Jr. 2000. "Introduction to Oliver C. Cox." In *Race: A Study in Social Dynamics*, ed. Oliver C. Cox. New York: Monthly Review Press.
- Reed, Adolph, Jr. 2002. "Unraveling the Relation of Race and Class in American Politics." *Political Power and Social Theory* 15: 265-74.
- Robin, Corey. 2004. *Fear: The History of a Political Idea*. New York: Oxford University Press.
- Roediger, David R. 1991. *The Wages of Whiteness: Race and the Making of the American Working Class*. New York: Verso.
- Rogin, Michael P. 1988. *Ronald Reagan, the Movie, and Other Episodes of Political Demonology*. Berkeley: University of California Press.
- Said, Edward W. 1990. "Zionism from the Standpoint of Its Victims." In *Anatomy of Racism*, ed. David Theo Goldberg. Minneapolis: University of Minnesota Press.
- Sears, David O. 1988. "Symbolic Racism." In *Eliminating Racism: Profiles in Controversy*, ed. P. A. Katz and D. A. Taylor. New York: Plenum.
- Sears, David O. 1994. "Urban Rioting in Los Angeles: A Comparison of 1965 with 1992." In *The Los Angeles Riots*, ed. Mark Baldassare. Boulder: Westview Press.
- Sidanius, Jim, and Felicia Pratto. 1999. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. New York: Cambridge University Press.
- Skrentny, John D. 2002. *The Minority Rights Revolution*. Cambridge: Harvard University Press.
- Smith, Rogers M. 1997. *Civic Ideals: Conflicting Visions of Citizenship in U.S. History*. New Haven: Yale University Press.
- Sniderman, Paul, and Thomas Piazza. 1993. *The Scar of Race*. Cambridge: Harvard University Press.
- Takaki, Ronald. 2000. *Iron Cages: Race and Culture in 19th Century America*. New York: Oxford University Press.
- Tate, Katherine. 2004. "Political Incorporation and Critical Transformations of Black Public Opinion." *Du Bois Review* 1(2): 345-59.
- Walters, Ronald. 1988. *Black Presidential Politics in America: A Strategic Approach*. Albany: State University of New York Press.
- Walton, Hanes, Jr. 1997. *African American Power and Politics: The Political Context Variable*. New York: Columbia University Press.
- Warren, Dorian T. 2004. "A New Labor Movement for a New Century? The Incorporation and Representation of Marginalized Workers in U.S. Unions." Ph.D. diss. Yale University.
- Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton: Princeton University Press.
- Zimbardo, Philip. 1973. "The Mind is a Formable Jailer: A Pirandellian Prison." *New York Times Magazine*, April 8.

# Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure

B. PETER ROSENDORFF *University of Southern California*

*The increased "legalization" embodied in the revised Dispute Settlement Procedure (DSP) of the World Trade Organization (WTO) is shown to be an institutional innovation that increases the opportunities for states to temporarily suspend their obligations in periods of unexpected, but heightened, domestic political pressure for protection. This increased flexibility in the system reduces per-period cooperation among states but also reduces the possibility that the regime may break down entirely. There is shown to be a trade-off between rigidity and stability in international institutional design in the face of unforeseen, but occasionally intense, domestic political pressure. In a model with a WTO that serves both an informational and adjudicatory role, it is established that agreements with DSPs are self-enforcing, are more stable, and are more acceptable to a wider variety of countries than agreements without DSPs. Evidence drawn from data on preferential trading agreements supports the key hypotheses.*

The world trading system has become significantly more "legalized" in the recent period (Goldstein and Martin 2000), with the adoption of the Dispute Settlement Procedure (DSP) as part of agreements forming the World Trade Organization (WTO). In contrast with national law, however, the WTO has no enforcement powers, "no jailhouse, no bail bondsmen, no blue helmets, no truncheons, no tear gas" (Bello 1996), to induce compliance. Absent any enforcement power, what function does international dispute settlement serve? Do these mechanisms condition state behavior in any significant way?<sup>1</sup>

The standard view of the WTO is that the institutions must be "capable of identifying and sanctioning (or at least authorizing sanctions against) cheating on the cooperative equilibrium" (Trebilcock and Howse 1999, 54). The hardening of the dispute provisions of the WTO is seen as making the potential punishments more severe, with the intent of extracting more cooperative behavior among the member states. The DSP also closes loopholes, eliminating "grey areas," limiting further the possibility of opportunistic behavior. The system is viewed now as *less* flexible, and this greater rigidity associated with the shift to legalization is expected to lead to more compliance (Goldstein et al. 2000; Smith 2000; Yarbrough and Yarbrough 1997).

B Peter Rosendorff is Associate Professor, International Relations and Economics, and Director, Center for International Studies, University of Southern California, University Park, Los Angeles, CA 90089-0037. Email: bpeter@usc.edu. Thanks to Bill Ethier, Todd Sandler, Helen Milner, seminar participants at the Regional Integration Network Meeting in Montevideo, 2004, American Political Science Association Annual Meeting in Philadelphia, 2003; at Public Choice Society Annual Meetings, San Diego, March 2002, Latin American and Caribbean Economic Association Annual Meetings, Montevideo, Uruguay, October 2001, to the editor and three anonymous referees.

<sup>1</sup> Beyond the DSP and the WTO, we seek to identify whether international law has an independent effect on the behavior of states. We might call this the "endogeneity problem": do states sign international agreements because they intend to comply anyway, or do they comply because they have signed an agreement and take actions (or refrain from taking actions) contrary to what they would have done absent the agreement (Chayes and Chayes 1993, Downs et al. 1996)?

Many scholars (and negotiators and WTO officials) have viewed the introduction of the WTO DSP as highly successful and effective (Jackson 1997a), and the relatively frequent appeal by its members to the (new, as of 1995) DSP is taken as evidence of the increased compliance with treaty provisions.

Others disagree. Reinhardt (1999) suggests that instead the frequent use of the DSP is not evidence of "success," but marks "potential challenges to the system" (2). The increased filings of disputes observed since the introduction of the revised DSP are not attempts to maintain cooperation but are instead evidence of increased violation of treaty obligations. Setear (1997) also argues that the enhanced DSP is a "step backward in the process towards greater cooperation": its relative ease of use increases opportunities for noncooperation and increases the likelihood of defection.

This article attempts to resolve this confusion and does so by clarifying the key role played by the DSP. The revised DSP enhances the stability of the cooperative regime; it does so not because it has become more rigid, but because it has become more *flexible*. The DSP now permits "compensation" for violations once authorized and emphasizes that the compensation is limited to an amount proportional to the loss experienced—and consequently adds a degree of flexibility which leads to an enhanced stability of the world trading system. Moreover, a wider variety of countries are more willing to sign an agreement with a DSP procedure of this type than any agreement without. Agreements with such a mechanism are easier to strike *ex ante* (cf. Fearon 1998).

Cooperation (at least in the long run) and discretion are therefore not mutually incompatible; it is overstating the case to argue that agreements have to be designed to deal with the "domestic political trade-off between treaty compliance and policy discretion" (Smith 2000, 138), or that "more" legalization may "threaten liberalization" (Goldstein and Martin 2000, 630). There is, instead, a trade-off between *rigidity* and *stability*. A DSP embodying the proportionality principle reduces

the rigidity of the system and increases the long-run stability of the cooperative institution.<sup>2</sup>

## DOMESTIC POLITICS

A country's negotiator at the international bargaining table is a political representative, responding (optimally) to and constrained by the political pressures it faces back home (Putnam 1988). Domestic political pressures and alignments, however, are subject to changes that are imperfectly anticipated or even unexpected. In periods in which the political pressure to provide some sort of protection to the domestic import-competing industry becomes unexpectedly acute, a government may be willing (in the absence of any means of escape) to abrogate its responsibilities under a trade agreement entirely to protect its domestic constituency (and its own incumbency). If, however, there are opportunities for signatories to escape their obligations (at least temporarily until the unexpected political pressure passes to a more "normal" state of affairs), an affected country may take such an opportunity while remaining within the parameters of the international agreement. One such avenue of escape is a willingness to be subject to the discipline of the DSP under the WTO. That is, a violation incurred for political reasons may be tolerated by other signatories under the agreement if the violation is temporary, and some sort of compensation scheme is available for the affected country(ies). The use of the DSP therefore allows a contracting partner to violate the agreement, compensate the losers, and still remain within the community of cooperating nations. Hence, an agreement with a DSP is less prone to abrogation by a state suffering intense political pressure to protect; such an agreement is more stable than one without a DSP.<sup>3</sup>

<sup>2</sup> A number of important works have argued the effectiveness of the General Agreement on Tariffs and Trade (GATT)/WTO system and its DSP. Bown (2001) claims that the DSP (or, more specifically, tolerated threats of retaliation) has been successful in generating liberalization. Staiger and Tabellini (1999) suggest that the GATT/WTO provides a (time-consistent) commitment device for governments in the game with their domestic political supporters. Bagwell and Staiger (1999) show that the principles of nondiscrimination and "most-favored nation," the cornerstones of the GATT/WTO system, lead to countries credibly forgoing beggar-thy-neighbor terms of trade shifts. This article like that of Ethier (2001), investigates the consequences of the limited punishment actions available and the lack of enforcement power.

<sup>3</sup> This argument also provides a justification for the existence of the escape clause and other safeguard measures of the GATT. Article XIX measures may play the role of providing a crucial escape valve for domestic political pressures that may have accumulated. See Rosendorff and Milner (2001) for a formal model that establishes that agreements with an escape clause Pareto dominate those without in the presence of imperfect information regarding future political pressures. Sykes (1991) provides a compelling argument that the purpose of Article XIX safeguards is to accommodate politicians' need to accommodate the pressures of materially injured sectors. A comparison of the DSP and Article XIX actions is discussed in what follows.

Although this paper studies the WTO's DSP in some detail, the underlying intuition regarding the utility of such procedures readily applies to other, regional agreements that embody dispute settlement procedures. See, for example, Gruber (1999) on supranational

A self-interested international negotiator pondering the gains and losses of entering into an international agreement may be more willing to sign such an agreement (and be constrained by its provisions) if s/he is aware that breach of its obligations is permitted under certain circumstances. The possibility that future political pressures to protect might become intense implies that an astute politician will want to preserve a policy instrument to deal with that pressure. Commitment to a trading regime where some sort of instrument remains in the hands of the politician is easier to achieve than without it. The "shadow of the future" stretches less far and is less penal when temporary accommodation to political pressure is available (Fearon, 1998; Rosendorff and Milner 2001; Sykes 1991). Whereas only the most patient politicians who value the future very highly can sustain cooperation in an environment without a DSP, an appropriately designed DSP can facilitate entry into the agreement by states less "patient," or with a lower valuation of the future.

## Enforcement and the Proportionality Principle

The self-enforcing nature of the agreement makes the DSP effective without explicit enforcement powers. An astute politician may prefer to protect a politically powerful industry in periods of unexpected stress and, at the same time, compensate its trading partners for any burden. Although the compensation demanded may be severe, the domestic political costs of paying the compensation are likely to be smaller than the political benefits from protecting the industry. But what is more crucial is that, by paying the compensation voluntarily, the country is signaling to its trading partners that it intends to return to a cooperative stance as soon as the temporary pressure rescinds. The payment of compensation acts as a signal of the country's intent to continue to cooperate in the future, an intent justified by the continued benefits of cooperation. The payment is a penalty paid to preserve a country's reputation as a cooperator (at least in "normal" times). In response, the trading partners observe this willingness to pay to preserve its reputation and opt not to punish the offending partner by revoking concessions or even exiting the system.<sup>4</sup>

governance in the North American Free Trade Agreement, or Busch (1999) on forum shopping across agreements, or Levy and Srinivasan (1996) on the effect of allowing private parties access to a regional agreement's DSP on the government's willingness to sign such an agreement.

<sup>4</sup> Reinhardt (2001) offers an explanation for the willingness of defendants to settle (offer a concession) prior to the determination of the DSP panel, absent enforcement. In a model where the defendant might be "compliant" and the plaintiff may be "tough," it may be cheaper for a compliant defendant to concede than to risk retaliation after a panel finding. Hence, the threat of retaliation makes the WTO process self-enforcing. Downs et al. (1996) argue that enforcement is not necessary—the WTO, members of which have self-selected themselves into the agreement, is fundamentally cooperative. Alternatively, enforcement is not necessary because the structure and rulings at the WTO reflect the underlying power relations of the

The “proportionality principle” (that compensation is limited to that which restores “balance” to previously negotiated concessions) is a crucial element of the DSP. If the cost associated with using the DSP was excessively large (the retaliatory punishment exceeds the political gains), countries would not be willing to apply these penalties to themselves and the DSP would lose its teeth. The proportionality principle limits the costs associated with adopting the DSP and thereby increases the stability of the system.

## CHARACTERIZATION OF THE DSP

The procedures specified in the Dispute Settlement Understanding (DSU) adopted during the Uruguay Round are consistent with the practice that had developed since the GATT was first implemented in 1947. A contracting party may file a complaint with the WTO regarding a perceived violation of the treaty on the part of another member. If formal, bilateral consultations are unproductive (an attempt at a negotiated resolution), the complainant may request that a panel of independent experts investigate the matter and make a recommendation (a more “judicial” approach). If the panel finds that the offending action is GATT-inconsistent, the offending party is obliged, should the panel so recommend, to terminate the violating measure and bring its practice back into conformity with its GATT obligations. The finding is “legally binding” on the members (Jackson 1997b) and can be appealed to the Standing Appellate Body, a panel of three experts drawn from a permanent roster of seven, selected for a four-year term on a staggered basis. There is no possibility that any member can “block” the report. If the recommendations of the panel are not implemented within a reasonable amount of time, the DSU permits possible “compensation” or retaliation. The purpose is to provide compensatory benefits to restore the balance of negotiated concessions disturbed by the noncomplying measure (Dunoff and Trachtman 1999; Jackson 1997b). If the offending state does not change its offending action or provide compensation, the WTO may authorize a retaliation to restore balance. Although the agreement clearly favors compliance with negotiated concessions, it is clear that the WTO system “authorizes a Member to choose to ‘breach’ an obligation, and pay compensation to the injured party” (Dunoff and Trachtman 1999, 26).<sup>5</sup>

disputants and, when retaliation is possible, compliance is observed (Garrett and Smith 2002).

<sup>5</sup> Jackson (1997a, 1998) argues that there is an obligation to “perform” under the terms of the agreement—an offending nation does not have a “choice” to compensate. Yet the DSU specifically authorizes retaliation if an offending country has not complied with a ruling. This view is consistent with the legal hypothesis of “Efficient Breach”—where breach is more efficient in a Pareto sense than is performance under a contract. In this view, the WTO can be viewed as an incomplete contract, and, while there is no true court-like mechanism to compel payment in the case of breach, here we show that voluntary compliance can work just as well. See Dunoff and Trachtman (1999).

In most cases, the defendants are found in violation.<sup>6</sup> In most of those, the defendants abide by the findings of the DSP. This is taken by a number of observers as evidence of the “success” of the institution. Punishments for breach of obligations under the treaty are usually set at a level “commensurate with the violations” (Ethier 2001) and only the country harmed is compensated (Jackson 1998).

The effect of a finding by a panel that there has been a violation is an obligation by the offending state to restore the losses experienced by the partner state. The DSP therefore takes four crucial actions: (1) it hears evidence of violation; (2) it rules whether or not a violation has occurred; (3) if a violation is identified, it estimates the compensation that is due; and (4) it reports that compensation has been made (by virtue of closing of the case).

The institution then serves a crucial information-providing role. It establishes the facts, adjudicates on a violation, estimates the damages, and reports a successful completion of the process. It is this informational role of the DSP that determines its effectiveness in the world trading system.<sup>7</sup>

## THE INSTITUTION AS EQUILIBRIUM

The DSP is a mechanism embodied within the broader set of institutions that govern trading relations between states. Following North (1990), an institution is viewed as an equilibrium to a game of strategic interaction. In what follows we specify a pair of strategies for two countries that embody a procedure for dealing with violations of a commitment to cooperate that is consistent with the dispute settlement procedures as articulated in the DSU. If this pair of strategies is a Nash equilibrium to the game of repeated strategic action that describes relations between trading states, then we can say that the DSP is an equilibrium institution. In the next section we compare two games of international trade policy played between two contracting parties. In the first game there is no DSP institution; in the second, at each period, the players listen to the information provided by the DSP. The institution with the DSP is shown to be more stable than without a DSP;

<sup>6</sup> Between 1973 and 1998, over 100 cases were paneled in which a defendant country has either raised its tariffs or refused to liberalize as agreed to. The defendant was found guilty in all but 9 of the cases (Bown 2001).

<sup>7</sup> Consistent with Keohane (1984), this explanation of the effectiveness of the DSP lies in its informational role, thereby reducing transaction costs and increasing transparency. For the informational role of multilateral institutions see Oye (1986). This role for the DSP has also been suggested by Kovenock and Thursby (1994) in a model without domestic politics but with a set of “demons” who introduce random deviations from the cooperative regime. Maggi’s (1999) model has the WTO informing third parties of any observed violation of a country’s obligations. The effect is to facilitate enforcement efforts. Similarly, Ozden (2001) has the DSP informing third parties if noncontractible implementing investments have been made. Here the information provided by the WTO prevents mistaken punishments from being applied. More generally, a variety of economic and political institutions have developed to provide crucial information to interacting parties; for example, the Law Merchant (Milgrom et al. 1990) or lobbies (Milner and Rosendorff 1996).



moreover, a wider variety of states will be willing to sign agreements when a DSP is available than when no such possibility exists.

## THE ECONOMY

Consider two countries that are identical, except for their endowments. Each country is endowed with and consumes three goods labeled  $x$ ,  $m$ , and  $z$ , where  $z$  is the numeraire good (with units chosen that the price of a unit of  $z$  is 1). On the supply side, the home country is relatively well endowed with good  $x$ , and the foreign country with  $m$ . More specifically, the world endowment of  $x$  and  $m$  are both set at unity; home is endowed with fraction  $\beta > \frac{1}{2}$  of good  $x$  and  $1 - \beta$  of  $y$ ; foreign is endowed with  $1 - \beta$  of  $x$  and  $\beta$  of  $y$ . The Heckscher–Ohlin theorem implies that home will import  $m$  and export  $x$ .

On the demand side, utility is assumed to be additively separable.  $U(x, m) = u(x) + u(m) + z$ . Each country has a single instrument at its disposal: a tariff on its imported good. Home can apply the specific tariff  $t$  on the imports of good  $m$  while foreign levies  $\tau$  on its imports of  $x$ . Utility maximization and market clearing implies that the price of  $m$  at home rises, and hence the home consumer surplus falls with  $t$ , whereas an increase in the tariff abroad actually lowers the price of  $x$  at home, raising consumer surplus. The domestic firms earn profits  $\Pi_m(t)$  (for the import-competing firms) and  $\Pi_x(\tau)$  for the export firms, with  $\Pi_m(t)$  rising with  $t$  and  $\Pi_x(\tau)$  falling with  $\tau$ . Tariff revenues are denoted  $T(t)$  which rise and then fall with  $t$ . The foreign country's payoffs are symmetric. A government's (one-period) utility depends on the sum of consumer and producer surpluses, and tariff revenues.<sup>8</sup> Moreover, political pressure, which import-competing firms bring to bear, is added to the objective function by weighting the firms' profit term. Let  $a > 0$  denote the weight that government attaches to firm's profits. The home government's (one-period) utility function then is  $G(t, \tau; a) = CS(t, \tau) + a\Pi_m(t) + \Pi_x(\tau) + T(t)$ . Similarly for the foreign government,  $G^*(t, \tau; \alpha) = CS^*(t, \tau) + \alpha\Pi_x^*(\tau) + \Pi_m^*(t) + T^*(\tau)$ , where  $\alpha$  is the weight put on the interests of the import-competing sector in foreign by the foreign government.<sup>9</sup>

The stochastic political pressure parameters  $a$  and  $\alpha$  are independently and identically distributed over the support  $(0, \infty)$ , with cumulative distribution function  $\Phi$ . At the beginning of each period, the government in each country knows the level of political pressure it faces at home; it is uninformed about the political pressures that have emerged in the foreign country.

<sup>8</sup> For any given level of the foreign tariff  $\tau$ , the home government's objective function rises and then falls with the home tariff  $t$ ; for any given  $t$ ,  $G$  falls with  $\tau$  since the marginal losses to the export firms always outweigh the benefits to the consumer of a higher foreign tariff.

<sup>9</sup> These "politically optimal objective functions" capture the idea that government officials are politically motivated (Baldwin 1987) and are consistent with the derived political support functions from a political contributions model such as those by Grossman and Helpman (1994).

And each is equally uninformed about the nature of the politics each might face in any future period.<sup>10</sup>

## International Cooperation

We characterize the fundamental problem of sustaining cooperation between countries in the realm of international trade.<sup>11</sup> Because we are interested in the role of the DSP within an ongoing agreement in which the tariff bindings have been previously set, we take the existence of a previously negotiated pair of cooperative tariffs  $(t^C, \tau^C)$  as given. Presumably they are the (Pareto optimal) pair of tariffs that maximize the present discounted value of the sum of both governments' expected utility over the infinite future. Moreover, since these were negotiated before the following games are played, these cooperative tariffs were set before the players are aware of the political conditions in the current period in their own countries.<sup>12</sup> At the beginning of each period, the players' types  $a$  and  $\alpha$  are revealed to each country—i.e., home sees  $a$  but not  $\alpha$  and foreign vice versa. Each country decides on its current period tariff rate simultaneously—whether to renege on the cooperative agreement (and apply the optimal defection tariff) or implement the cooperative tariff. This extensive form description yields payoffs that can be written in the normal form of a standard prisoner's dilemma (PD).

**Under Cooperation.** The utility of the home government under cooperation is  $G(t^C, \tau^C; a) = CS(t^C, \tau^C) + a\Pi_m(t^C) + \Pi_x(\tau^C) + T(t^C) \equiv C(a)$ , which is not a function of  $\alpha$ . Similarly,  $C^*(\alpha) = G^*(t^C, \tau^C; \alpha)$ . Since the cooperative agreement is negotiated before any details of the domestic politics in either country are revealed, the payoffs for each country are functions only of each country's domestic politics parameter.

**Under Nash Equilibrium.** Under the Nash equilibrium (NE) to the one-shot game, each player chooses a level of domestic trade barriers as a best response to the behavior of the opponent. In any period in which  $a$  and  $\alpha$  are known, we can solve for the NE in trade barriers for that period. Let  $t(\tau) = \arg \max_t G(t, \tau; a)$ , and,  $\tau(t) = \arg \max_\tau G^*(t, \tau; \alpha)$ , and, solving simultaneously, we obtain the Nash pair  $(t^N, \tau^N)$ . Denote home government's utility under the Nash as  $N(a, \alpha) = G(t^N(a, \alpha), \tau^N(a, \alpha); a)$ .

**What about Defections?** What are the payoffs when, say, home defects and foreign cooperates in the one-shot game? The optimal defection is  $t^D = \arg \max_t G(t, \tau^C; a)$ , and utility under the optimal defection is  $D(a) = G(t^D(a), \tau^C; a)$ . If, instead, foreign

<sup>10</sup> This one-sided asymmetry of information is in fact not necessary for the results that follow but is useful to maintain for ease of exposition and seems the most realistic of the possibilities. What is key for the results is the presence of uncertainty regarding the political future either country might face in future periods.

<sup>11</sup> This section follows Rosendorff and Milner (2001), (RM).

<sup>12</sup> Hence,  $t^C$  and  $\tau^C$  are exogenous from the point of view of the current game and, therefore, not functions of the political parameters drawn for the current period,  $a$  and  $\alpha$ .

defects and home cooperates, home receives the sucker's payoff:

$$S(a, \alpha) = G(\tau^C, \tau^D(\alpha); a).$$

### The Prisoners' Dilemma

In any play of the game we have  $D(a) > C(a) > N(a, \alpha) > S(a, \alpha)$  for any pair  $(a, \alpha)$ , a PD, as represented by the standard  $2 \times 2$  normal form matrix.

	$C^*$	$D^*$
$C$	$C(a), C^*(\alpha)$	$S(a, \alpha), D^*(\alpha)$
$D$	$D(a), S^*(a, \alpha)$	$N(a, \alpha), N^*(a, \alpha)$

To simplify the notation,  $D(a) - C(a) \equiv B(a)$ . Notice that, as the political pressure to protect becomes larger, the gains from defection relative to cooperation rise; i.e.,  $B'(a) > 0$ . Each player is susceptible to political pressure both to protect against foreign imports and to open export markets; in the future both are equally unsure how much pressure each will experience.

This one-shot game is infinitely repeated, and the players choose strategies to maximize the expected sum of their discounted one period utilities:

$$E \sum_{t=0}^{\infty} \delta^t G(t, \tau_i; a_t) \quad \text{and} \quad E \sum_{t=0}^{\infty} \delta^t G^*(t, \tau_i; \alpha_t)$$

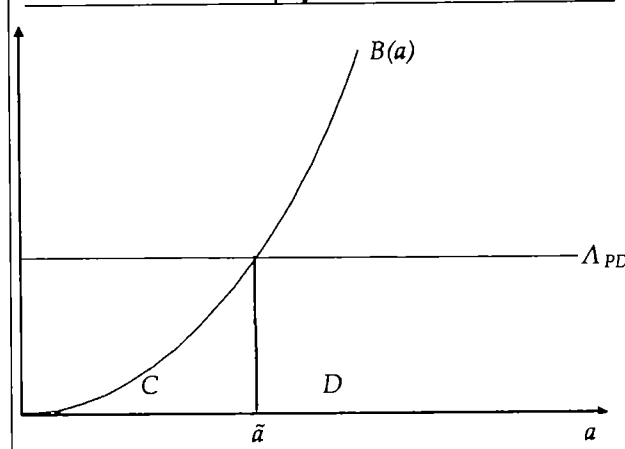
We can view the cooperative outcome to this game as characterizing the international agreement in the absence of a mechanism to deal with disputes or other episodes of unfair practice. We consider an NE supported by the usual grim trigger—an infinite punishment in the event of a deviation. In equilibrium, each player cooperates every period until the domestic political shock breaches some threshold. Then defection occurs and punishment continues henceforth.

**LEMMA 1.** *The following pair of strategies constitutes an equilibrium: for some  $\tilde{a}$  in the support of  $a$ , home plays  $C$  if  $a < \tilde{a}$ , or plays  $D$  if  $a > \tilde{a}$  or if  $D^*$  has been played by foreign in the past; for some  $\tilde{\alpha}$  in the support of  $\alpha$ , foreign plays  $C^*$  if  $\alpha < \tilde{\alpha}$ , or plays  $D^*$  if  $\alpha > \tilde{\alpha}$  or if  $D$  has been played by home in the past.*

The proofs are in the Appendix. The incentive to defect in any period with draw  $a$  is  $B(a)$ . If the incentive to defect is less than the present discounted expected losses of future punishments, cooperation is sustained; i.e., the no-defect condition is  $B(a) \leq \frac{\delta}{1-\delta}(C - N) \equiv \Lambda_{PD}$ . Define  $\tilde{a}$ , such that  $B(\tilde{a}) = \frac{\delta}{1-\delta}(C - N)$ , and the no-defect condition becomes simply  $a \leq \tilde{a}$  (since  $B'(a) > 0$ ). In this equilibrium, each player cooperates until the pressure to protect gets extraordinarily high. Then it defects and incurs the infinite punishment.<sup>13</sup> This equilibrium is illustrated in Figure 1.

<sup>13</sup> If there is an upper bound to the magnitude of the shock, say  $a^{\max}$ , then, for all discount factors large enough, cooperation dominates defection. That is, cooperation is assured in any period iff  $B(a^{\max}) \leq \frac{\delta}{1-\delta}(C - N)$ , or iff  $\delta > \frac{B(a^{\max})}{C - N + B(a^{\max})}$ . Therefore, we have the standard result that cooperation (here in the face of political pressure) is sustainable only when the players are very patient (when maximal shocks are finite).

**FIGURE 1.** The No-Defect Condition for the PD Game Is  $B(a) \leq \frac{\delta}{1-\delta}(C - N) = \Lambda_{PD}$



### The Dispute Settlement (DS) Game

Consider a period in which the government of the domestic country receives a high shock in the PD game. The unexpectedly large value of  $a$  implies that the government has come under excessive pressure to protect the local import-competing industry. The options that the government faces are first to stick with the purely cooperative agreement and play  $\tau^C$  and earn  $C(a)$ , or alternatively to play  $\tau^D(a)$  and earn  $D(a)$ . In the event that home chooses to play the defect strategy,  $D$ , it would invoke an infinite punishment in the PD game, and the cooperative world trading system breaks down.

**The Ruling.** Instead consider the following DS game structure. After the domestic country has adopted a defect action (a violation) in any period, the other country (the complainant) files a dispute (requests a panel) with the WTO. Then the panel hears the case and makes a decision. If it finds a violation, it also decides on a penalty; finally the defendant decides whether to pay the penalty or not. The panel faces the same informational constraints as the other players. If the foreign country has played  $\tau^C$ , no information can be gleaned about foreign's domestic politics; however, home has defected and played  $\tau^D(a)$ , both foreign and the panel can invert the function describing the optimal defect tariff, and infer the state of politics at home,  $a$ . When  $a$  is known to all the players we will designate it  $\hat{a}$ . Clearly, both players and the DSP identify that a defection has occurred, since all can see that  $\tau^D(a) > \tau^C$ . But the WTO permits a country to rescind its commitments in various instances. For instance, the defendant might argue that it has become concerned that the good is not safe for human, plant, or animal health, or that its continued import may harm the environment (such as the debate over genetically modified foodstuffs between the United States and the European Union (EU), or hormone-fed beef). The DSP will have to make a determination as to whether the measure is "a disguised restriction on international trade" for political purposes or a legitimate health,

safety, or environmental measure<sup>14</sup>. The probability that the DSP finds in favor of the plaintiff is set (as in Reinhardt 2001) at  $\theta \in (0, 1)$ , which is common knowledge.

Should the panel find in favor of the plaintiff, the panel will attempt to measure the loss that the complainant has sustained due to the defect action of the other member. If home defects, foreign receives  $S^*(a, \alpha)$  instead of  $C^*(\alpha)$ . Then the actual losses are  $L^*(a, \alpha) = C^*(\alpha) - S^*(a, \alpha)$ . However, the panel cannot verify (ex ante) the actual value  $\alpha$ . Instead it must take its best guess given the information at its disposal. The actual estimate of the losses incurred by foreign will be the expected value of  $L^*(\hat{a}, \alpha)$ , since  $\hat{a}$  is known, which we write as  $L^*(\hat{a}) = \int_{\alpha} L^*(\hat{a}, \alpha) d\Phi$ . Similarly, if foreign is the defecting party, the panel will establish a compensation of  $L(\hat{a}) = \int_{\alpha} L(a, \hat{a}) d\Phi$ , which is the expected losses experienced by home. As far as the plaintiff country is concerned, the expected loss (before the ruling is made) for which it will be compensated is  $\theta L^*(\hat{a})$  for foreign and  $\theta L(\hat{a})$  for home. For notational purposes, define  $L = \int_{\alpha} \int_{\alpha} L(a, \alpha) d\Phi d\Phi$ .

**Compensation.** In any period of the DS game, a player can take the Pareto action, i.e., play  $C$  as in the aforementioned PD, or can incur the costs of violating the agreement and pay the compensation,  $DS$ , at expected cost  $\theta L(\hat{a})$  (or  $\theta L^*(\hat{a})$ , which accrues to the other player) or can defect  $D$  as before. That is, the panel has no enforcement powers, and the defendant can choose not to pay any compensation whatsoever. Any compensation that is paid in equilibrium is done so on the volition only of the defendant state.<sup>15</sup>

The stage game payoffs can now be expressed in the normal form as a  $3 \times 3$  matrix, with the payoff structure as in the table below, viewed from the moment the political pressures are revealed but before any (possible) ruling by the DSP panel:

	$C^*$	$DS^*$	$D^*$
$C$	$C(a), C^*(\alpha)$	$S(a, \hat{a}) + \theta L(\hat{a}), D^*(\hat{a}) - \theta L(\hat{a})$	$S(a, \hat{a}), D^*(\hat{a})$
$DS$	$D(\hat{a}) - \theta L^*(\hat{a}), S^*(\hat{a}, \alpha) + \theta L^*(\hat{a})$	$N(\hat{a}, \hat{a}) + \theta L(\hat{a}) - \theta L^*(\hat{a}), N^*(\hat{a}, \hat{a}) + \theta L^*(\hat{a}) - \theta L(\hat{a})$	$N(\hat{a}, \hat{a}) - \theta L^*(\hat{a}), N^*(\hat{a}, \hat{a}) + \theta L^*(\hat{a})$
$D$	$D(\hat{a}), S^*(\hat{a}, \alpha)$	$N(\hat{a}, \hat{a}) + \theta L(\hat{a}), N^*(\hat{a}, \hat{a}) - \theta L(\hat{a})$	$N(\hat{a}, \hat{a}), N^*(\hat{a}, \hat{a})$

To describe the NE to this game, it is necessary to divide the support of the politics parameter  $a$  into three subsets. Define  $\underline{a}$  such that  $\theta L(\underline{a}) = B(\underline{a})$ ; define  $\bar{a}$  such that  $\theta L(\bar{a}) = \frac{1}{1-\theta}(p^2(N - S - D + C) +$

$p(D - 2N + S)) \equiv \Lambda_{DS}$  (where  $p = Pr(a < \underline{a})$ , and  $I = \int_{\alpha} \int_{\alpha} I(a, \alpha) d\Phi d\Phi$  for  $I = D, N, S, C$ ). Lemma 3 in the Appendix shows that a sufficient condition for  $\underline{a} < \bar{a}$  is  $\theta < \frac{\Lambda_{DS}}{L(\underline{a})}$ .

**Definition 1.** A Dispute Settlement Strategy (for home), denoted DSS, is a strategy in which home, having drawn politics of type  $\hat{a}$ , plays  $D$  if  $D^*$  has been played in any period in the past; otherwise home plays  $C$  if  $\hat{a} < \underline{a}$ , plays  $DS$  if  $\underline{a} \leq \hat{a} \leq \bar{a}$ , and plays  $D$  if  $\hat{a} > \bar{a}$ .

As before, the extent of the political shock determines the gains to be had from defection this period; these gains rise with the political pressure that the firms bring to bear; i.e.,  $B'(a) > 0$ . It is also the case that the (expected) losses experienced by one's trading partner rise as  $a$  rises; i.e.,  $\theta L'(a) > 0$ . Given a draw  $\hat{a}$ , if the gains from defection are small relative to the compensation it would have to pay ( $B(\hat{a}) < \theta L(\hat{a})$ , i.e.,  $\hat{a} < \underline{a}$ ), then the government sticks to the Pareto play  $C$ . If the expected penalty is not too onerous ( $\theta L(\hat{a}) < \Lambda_{DS}$ , or  $\hat{a} \leq \bar{a}$ ), then the possible gains from defection ( $\theta L(\hat{a}) \leq B(\hat{a})$ , which occur when  $\underline{a} \leq \hat{a}$ ) cause the government to violate its obligations and suffer the consequences of a negative finding by a WTO panel,  $DS$ , which includes the (expected) payment of compensation. If the gains from defection are large, and the penalty is too large ( $B(\hat{a}) > \theta L(\hat{a}) > \Lambda_{DS}$ , or when  $\hat{a} > \bar{a}$ ), then the government ceases to cooperate entirely, violates its treaty obligations, and refuses to pay any compensation. A useful way to summarize the government's strategy is to say that the government cooperates (by playing  $C$  or  $DS$ ) when  $\min(B(\hat{a}), \theta L(\hat{a})) \leq \Lambda_{DS}$  and defects otherwise (see Figure 2).

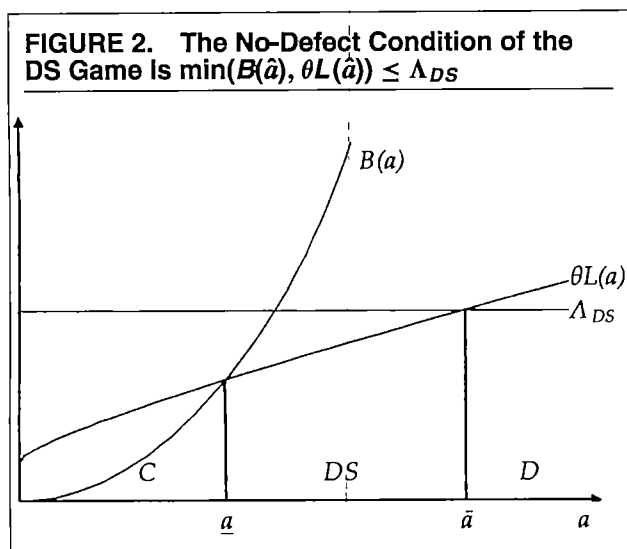
<sup>14</sup> Article XX of the WTO agreement.

<sup>15</sup> Note that the cost associated with "escape" here—the use of the DSP mechanism—is endogenous, and changes period by period. Moreover, the event that the cost is actually applied to the defendant is stochastic (it occurs with probability  $\theta$ ) and is the outcome of the dispute settlement process. This is a generalization of the study of escape clauses by Rosendorff and Milner (2001), in which the cost of escape was exogenous to the repeated game, in which the escape cost is incurred with certainty if the state chooses the escape clause action in any period, and in which the WTO had no arbitration role and merely reported if the offending country has penalized itself by incurring some exogenous adjustment costs.

The critical value of  $\Lambda_{DS}$  is the level of the cost such that, if the government plays the "cooperate" strategy (either  $C$  or  $DS$ ) into the indefinite future, the expected (net) benefits from doing so are equal to the expected benefits of defecting once and exiting the system forever. It is intuitive, therefore, that if the costs of the dispute procedure and the gains from defection are large, the government will cease to cooperate entirely.

**Proposition 1.** A pair of DSS strategies is an NE.

An international agreement with features similar to the DSP emerges as an equilibrium to the game of international trade. In any period a country (say, home) can stick to the cooperative deal and play  $C$ , or in response to political pressure  $\hat{a}$ , can defect to  $D$  ( $\hat{a}$ ). The other



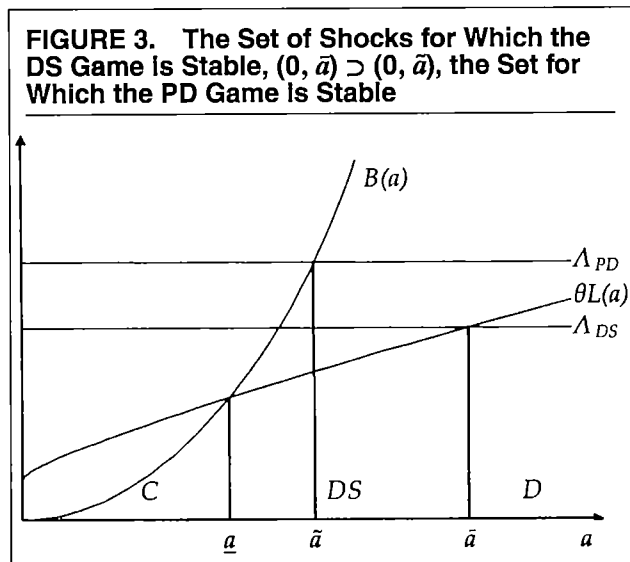
country (foreign) observing that the tariff applied is larger than the agreement tariff ( $t^D(\bar{a}) > t^C$ ) files a complaint. The panel also observes  $t^D(\bar{a})$  and undertakes its first responsibility, to adjudicate if the violation of the tariff binding is permissible under WTO codes. If it finds against the defendant, the panel must now try to establish its best guess of the loss foreign has incurred,  $L^*(\bar{a})$ . This penalty is paid voluntarily; the WTO verifies the payment and closes the file. Ex ante—at the time of the decision to violate—the expected cost of the violation is  $\theta L^*(\bar{a})$ . Cooperation resumes in the next period. Hence, the role of the institution is to verify the facts of the matter to its best ability, make a legal ruling, and then to rely on the voluntary behavior of the participants.<sup>16</sup>

The next proposition and its corollary establish the central results of this paper: agreements with a DSP are more stable than those without. We establish that the set of shocks that the agreement can withstand is greater when a DSP is present; i.e.,  $\bar{a} < \bar{a}$ . The corollary establishes that it is exactly those countries who are not patient enough to sustain cooperation in the pure PD who will gain from incorporating a DSP into the agreement. The DSP effectively lowers the threshold value of the discount rate necessary to sustain a cooperative outcome.

**Proposition 2.** *The DSP is more stable than the PD; i.e., the per-period probability of breakdown is smaller under the DSP than under the PD.*

The set of shocks that can be withstood without the equilibrium breaking down under the DSP,  $(0, \bar{a})$ , is

<sup>16</sup> Downs and Rocke (1995) present a series of games of international cooperation in the face of uncertain domestic politics, not unrelated to the game presented here. They argue that less severe punishments are necessary than the grim trigger required here in order to facilitate cooperation or, alternatively, a probabilistic approach to punishment. Agreements must therefore incorporate a degree of “optimal imperfection” to be effective. Here we include the dispute settlement strategy in the action space and obtain long-run cooperation under the grim trigger, and without uncertainty about whether the punishment, once authorized, will be applied



larger than (is a superset of) the set of shocks that can be withstood in the PD game,  $(0, \bar{a})$ . The implication here is that the DSP game is more robust against political shocks than is the pure PD version. To prove this result, we superimpose Figures 1 and 2 in Figure 3. Notice that  $\Lambda_{DS} = \frac{\delta}{1-\delta}(p^2(N-S-D+C) + p(D-2N+S)) \leq \frac{\delta}{1-\delta}(C-N) \equiv \Lambda_{PD}$  for all  $p \in [0, 1]$ . This can be seen by considering the PD game as a special case of the DS game where  $p = 1$ , then  $\Lambda_{DS} = \Lambda_{PD}$ ; when  $p = 0$ ,  $\Lambda_{DS} < \Lambda_{PD}$ , and  $\Lambda_{DS}$  is monotonic in  $p$ . As the diagram is drawn, it is easily observed that  $\bar{a} \leq \bar{a}$ ; interestingly, for the result to hold in general, we require  $\theta \leq \min\{\frac{\Lambda_{DS}}{L(\bar{a})}, \frac{\Lambda_{PD}}{L(\bar{a})}\}$ ; i.e., there must be sufficient uncertainty about the decision of the DSP. The effect of this restriction is to require (weakly) that there is an upper bound on the costliness of making use of the DSP. In addition to limiting the loss to an estimate of the damages incurred, the loss is lower in expectation if there is some probability that the panel will not penalize the offending member state.

**Corollary 1.** *Stable agreements with a DSP are feasible for a wider variety of countries than one without.*

If the variety of countries is captured by variations in their discount factors, a long-term stable equilibrium (one that does not break down) in the PD game is feasible only if both countries have discount rates that are high enough. For the PD game, the equilibrium is stable if and only if  $\delta > \frac{B(a^{\max})}{\Lambda_{PD} + B(a^{\max})}$ ; for the DS game, the equivalent condition is  $\delta > \frac{\theta L(a^{\max})}{\Lambda_{DS} + \theta L(a^{\max})}$ ; since  $\theta \leq \frac{\Lambda_{PD}}{L(\bar{a})}$  (the condition for the existence of the equilibrium), then  $\frac{\theta L(a^{\max})}{\Lambda_{DS} + \theta L(a^{\max})} < \frac{B(a^{\max})}{\Lambda_{PD} + B(a^{\max})}$  and, hence, a larger set of (i.e., lower) discount factors can support a stable equilibrium under a PD. Countries with low discount factors which might not have been able to join a stable PD agreement are now able to join a stable agreement by virtue of the DSP.

## The Trade-off between Rigidity and Stability

This extra stability of the DS game comes at a price, of course. Consider the interval  $(\underline{a}, \bar{a})$  in Figure 3. In the PD game, we would see pure cooperation; in the DS game for a shock in this interval, DS is played. The DS action is a defect action. There is no “true” cooperation in that period. Therefore the extra flexibility provided by the DS action (permitting cooperation when it was previously not possible) comes at the cost of its being used in periods when pure cooperation was previously available. Hence, an agreement with a DSP yields lower per-period cooperation (less rigid) but has a lower probability of breakdown (more stable).<sup>17</sup>

## THE PRICE OF ESCAPE

A similar argument provides a rationale for the inclusion of escape clauses in the WTO agreements (Rosendorff and Milner 2001; Sykes 1991).<sup>18</sup> Article XIX allows signatories to renege on their commitments under certain circumstances. When increased imports “cause or threaten serious injury to domestic producers” of import-competing goods, a country may, for a limited time, suspend its obligations under the GATT/WTO. This clause allows governments to escape their commitments in periods in which domestic industries are under pressure from increased imports from abroad.<sup>19</sup>

There are other forms of escape available throughout the GATT. Article VI of the GATT, the Antidumping (AD) and Countervailing Duties (CVD) codes (all part of the GATT agreement), allow member states to apply duties when imports are “dumped” or “sold at less than fair value” or when the foreign competitor is being subsidized (in the case of CVD), and these have the same effect of allowing temporary relief when the local industry comes under pressure from foreign competitors and/or increases its lobbying and political pressure on the local government. Balance of payments exceptions (Articles XVII and XII), infant industry protection (XVIII), and tariff renegotiation (XXVII)

<sup>17</sup> One can estimate the price of this extra stability: consider the set of shocks  $(\underline{a}, \bar{a})$ —in the PD, we would see pure cooperation, in the DS game we see use of the DSP. The expected loss from the DSP relative to the PD is  $\int_{\underline{a}}^{\bar{a}} [C(a) - (D(a) - \theta L^*(a))] d\Phi$ . The expected gains accrue when the shocks lie in the interval  $(\bar{a}, \bar{a})$ :

$$\int_{\bar{a}}^{\bar{a}} [D(a) - (D(a) - \theta L^*(a))] d\Phi = \theta \int_{\bar{a}}^{\bar{a}} L^*(a) d\Phi$$

<sup>18</sup> A crucial distinction between the model here and that of RM is that in this model we require direct compensation of the injured trading partner, whereas in RM the offending state simply penalized itself by incurring some adjustment costs. Since the foreign country was not receiving a payment, its willingness to tolerate a temporary defection is actually lower in an escape clause environment than in one with a DSP. Hence, an escape clause can sustain cooperation only if the cost of doing so is higher, *ceteris paribus*, than the cost of exercising the DSP.

<sup>19</sup> It is sometimes called a “safeguard” action.

all allow temporary escape from a country’s obligations under the GATT. Optimal institutional design is to include possibilities for escape or relief when unanticipated political pressures become too intense to endure without some sort of accommodation.<sup>20</sup>

The agreement has over time made escape more accessible, or easier to achieve; as a consequence we have seen an increased use of these measures. Some scholars and a number of negotiators have argued that it is time to tighten up some of these practices—an attempt to reform the Antidumping practices was unsuccessful during the Uruguay Round. The question is effectively: How easy should it be for a state to obtain tolerated relief? The model provides a clear way to think about this: lower costs clearly mean more frequent, tolerated escape, and less per-period cooperation. But it also works to increase the stability of the agreement and may permit more countries to accede to the deal. Stricter rules mean more cooperation, but fewer members and a more unstable agreement.

## Evidence

In 1996, the United States requested a DSP panel arguing that the EU’s prohibition on the imports of beef treated with hormones was inconsistent with its obligations under the WTO. The panel found that the EU ban was unjustified on a number of grounds, and the decision was upheld by the appellate body. Arbitration resulted in agreement that 15 months would elapse by which time the ban was to be removed. The EU did not comply with the finding and failed to remove the offending measure within that time period. The panel authorized retaliation/compensation of \$116.8 million (and C\$11.3 million in a similar case filed by the Canadians). The EU has not complied, and the United States continues to suspend concessions. Hoekman and Kostecki (2001) remark that the EU was politically unable to comply with the ruling: “Political constraints reflecting a strong lobby in the EU that opposed the use of hormones in meat production made it (compliance) impossible” (84). In addition, any increase in the productivity of European beef farmers would actually increase the costs of the common agricultural policy, something the EU could ill afford.

Other cases fit this pattern—a panel ruling to cease the offending measure, with which the defendant fails

<sup>20</sup> Hoekman and Kostecki (2001) describe these exceptions as “safety valves” (38), designed specifically to deal with political and social problems associated with increased imports. Sykes (1991) suggests that political gains to one party of exercising an escape clause must be larger than the losses that accrue to the trading partner for an escape clause to be “politically Pareto efficient.” Notice we make no such demand here—rather the payment of the penalty acts as a signaling device of the intention of the rogue state to return to the fold of cooperating nations.

Notice that RM, Sykes (1991), and this paper all require some penalty to be paid for demanding relief that is tolerated by the trading partners. In that sense, these opportunities for escape resemble the penalties a private contractor might incur if it chose to breach a contract. Such a promisor might find it preferable to renegotiate, or pay damages, once the time to perform arrives rather than perform under the terms of the contract.

to comply because the political costs of doing so are too high, and retaliation is authorized and applied: the Bananas case, in which the United States retaliated (with DSP authorization) against the EU in 1999 by applying tariffs up to 100% on a rotating set of goods, valued at \$191 million annually. This remained in effect until July 2001. In February 2000, a DSP panel found that Australia was illegally subsidizing the manufacture of automotive leather and was ordered to cease the measure and reimburse about \$19 million to the plaintiff, the United States. Perhaps in retaliation to the outcome in the Bananas case, the EU filed a dispute over a U.S. tax rule that allowed U.S. exporters to establish an offshore Foreign Sales Corporation. The panel authorized, in August 2002, penalties of 100% tariffs on \$4 billion worth of trade and raises, according to Lawrence (2003), the average import-weighted tariffs on U.S. exports to the EU by 1.8%, enough to wipe out the gains made during the Uruguay Round.

A similar dispute concerned the 30% U.S. steel tariffs applied in 2002. As a consequence of the close presidential election of 2000, the steel industry in electorally pivotal states like West Virginia, Ohio, and Pennsylvania were able to apply increased political pressure and extract temporary protection under the safeguard provisions. The appellate body of the DSP ruled the tariffs illegal and authorized compensation to the plaintiffs (EU, China, South Korea, Brazil, Switzerland, Japan, New Zealand, and Norway) of \$2.2 billion.

As to the durability of regimes with DSPs, we can look to the recent proliferation of regional and preferential trading arrangements, many of which have adopted dispute resolution mechanisms of various kinds. These mechanisms vary from "soft"—ad hoc negotiations among the parties to "hard"—standing independent tribunals whose determinations are legally binding. Smith (2000), for instance, examines a set of 63 post-1957 Preferential Trading Agreements (PTAs) and explains variations in the degree of "legalism" or "bindingness" of the DSPs by the degree of economic asymmetry of the signatories, especially when interacted with the proposed depth of liberalization.

While the argument here has focused on the institutions of the WTO, a similar logic applies to any PTA with a DSP with the aforementioned characteristics. Using the richer universe of PTAs, we can consider two testable hypotheses emerging from the model:

1. Those PTAs with DSPs, especially those that embody the proportionality principle, will be more durable, or last longer, than those without such an institutional characteristic.
2. The number of signatories will be higher in those agreements that embody a DSP relative to those that do not.

Pevehouse et al. (2002) estimate a duration model of PTA survival and find that those that embody a DSP have a lower failure rate. Using a sample of 85 agreements, they show, *ceteris paribus*, that the presence

of a DSP is positively and significantly related to the duration of the PTA. Moreover, using Smith's (2000) ranking of the degree of "legalization" of the DSP, Pevehouse et al. (2002) find that more legalism results in longer-lasting agreements.<sup>21</sup>

We can use the Smith's (2000) data to investigate whether the number of signatories rises with a DSP. Smith's "legalism" is measured on a five point scale: is third-party review in the instance of disputes available (0, if not)? If available, is the determination of the review panel binding (coded 1 if not)? If binding, is there a standing tribunal of judges (2 if not)? And is a standing panel present, but only states have standing before it (coded 3)? If states, treaty organs, and individuals can bring complaints, the degree of legalism is coded 4. Although this measure maps incompletely to the question at hand (is a DSP present and does it embody the proportionality principle) the correlation is likely to be very close.

The correlation between the number of members in an agreement and the Smith (2000) measure of legalism for the agreement is positive (0.27) and significantly different from zero at the 5% level ( $p = 0.034$ ). Higher levels of legalism are associated with larger numbers of signatories. Of course, this correlation is merely suggestive but does lend some support to the hypothesis.<sup>22</sup>

Anecdotal evidence as to the EU's bargaining position during the Uruguay Round also lends support to the argument that key to cooperation in the presence of political uncertainty requires retaliation, if authorized, to be limited in magnitude to an estimate of the losses incurred. Section 301 of the U.S. Trade Act of 1974 gave the President the authority to unilaterally retaliate against its trading partners if their practices were deemed by the President to restrict U.S. exports. This law was rendered more ominous by changes in 1988 (Super 301), which required the U.S. Trade representative to identify targets and set dates for retaliation. There was no limit to the number of countries, or the value of the punishments. During the Uruguay Round negotiations, the EU Commission frequently expressed its concern about the provisions in U.S. law for unilateral behavior and saw the revised DSP as a measure to bind the United States to the same legal standard as other members. The United States (and other WTO members) is precluded from making unilateral determinations as to violations, or nullification or impairment of benefits, and members must appeal to the DSP. The room for unilateral measures has shrunk.

<sup>21</sup> Other explanatory variables include measures of democracy, economic concentration, and depth of integration, and they control for GDP, the presence of a major power in the region, the number of members, the presence of military disputes and the degree of hegemony in the international system. Democracy and the DSP variables are consistently significant. See Mansfield et al. (2000, 2002) on the links between regime type and trade policy.

<sup>22</sup> A referee correctly remarks that there is the possibility of selection bias here: members may sign on to the PTAs with good DSPs because they have more harmonious relations, perhaps. Nevertheless, the link between durability and the number of signatories to PTAs and the presence of a DSP is striking.



The EU was clearly more inclined to sign on to the new WTO as a result of the removal of the possibility of highly punitive, unilateral trade sanctioning measures (Bhagwati and Patrick 1990).

### DSP and Cooperation

Whereas the WTO is a multilateral system, the analysis here is bilateral—two countries engaged in a dispute. Maggi (1999) establishes the supremacy on welfare grounds of a system of (unlimited) punishments from multiple sources relative to bilateral punishment (especially when small countries are involved); this paper establishes the benefits of a system of limited punishments—limited by the level of punishment from a single source for a temporary period. Although the approach here is clearly limited to circumstances in which countries have the capacity to inflict nontrivial harm on each other, extending this approach to a multilateral context may yield further insights that are relevant to small countries. However, the WTO's DSP explicitly limits standing (and restricts any authority to retaliate) to those member states whose concessions have been "nullified and impaired" by an offending state's actions.<sup>23</sup>

We might reevaluate the debate mentioned in the introduction. Does the introduction of the DSP increase or reduce the degree of cooperation between states who are signatories to the WTO? The answer is that it does both. If we have two countries that are very patient, then they always cooperate irrespective of the domestic political conditions, and adding a DSP will reduce the per-period cooperation by allowing temporary defections. If, however, the signatories are of moderate patience (where we would expect most countries to fall), then the agreement (before a DSP) carried the risk that at some point a political shock would hit that is large enough to warrant complete abrogation of the treaty and an exit from the system. Such an agreement generates much cooperation while in place but runs the risk of breakdown, a probability one event at some point in the future. Hence, an introduction of a DSP reduces the per-period cooperation (i.e., some periods there is temporary, tolerated defection), but the risk of breakdown of the entire treaty falls.

So yes, there are more disputes and less cooperation at any instant; but the agreement is clearly more stable and better able to endure despite the vicissitudes of domestic politics that affect the willingness of the signatories to remain within the community of cooperating nations.

<sup>23</sup> The DSU permits third-party involvement in the DSP in some instances, when more than one member wishes to complain about the "same matter"—that is, the third parties are alleging harm from the original infraction (Article 9). Article 10 permits "interveners" to make oral or written submissions to a panel. Given the emphasis on "rebalancing" concessions, however, an unharmed country has no compensation due and cannot lawfully engage in "retaliation" (Trebilcock and Howse 1999).

## APPENDIX

**Definition 2.** Let  $N(a, \alpha) - S(a, \alpha) \equiv A(a, \alpha)$ .

Recall that  $D(a) - C(a) \equiv B(a)$ . Without loss of generality, normalize the cooperative tariff  $t^C = t^D(0)$ ; i.e.,  $B(0) = 0$ .

**Proof of Lemma 1.** The no-defect condition is  $B(a) \leq \frac{\delta}{1-\delta}(C - N)$ . Define  $\tilde{a}$ , such that  $B(\tilde{a}) = \frac{\delta}{1-\delta}(C - N)$ . The no-defect condition becomes  $a \leq \tilde{a}$  (since  $B'(a) > 0$ ). ■

**Lemma 2.** In the nondegenerate case, there exists  $\underline{a}$  such that  $B(\underline{a}) = \theta L(\underline{a})$ , and  $B(a) < \theta L(a)$  for  $a < \underline{a}$ , and  $B(a) > \theta L(a)$  for  $a > \underline{a}$ .

**Proof.**  $B'(a) = G_{t|D(a)} + \Pi_m(t^D(a)) - \Pi_m(t^C) > 0$ ;  $B''(a) = \frac{d\Pi_m(t^D(a))}{da} \frac{dt^D(a)}{da} > 0$ . So  $B$  intersects the origin, and rises at an increasing rate. By comparison,  $\frac{dL(a)}{da} = -\frac{d}{da}(CS^*(t^D(a), \tau^C) + \Pi_m^*(t^D(a))) > 0$ ; and  $\frac{d^2L(a)}{da^2} = -\frac{d}{da^2}(CS^*(t^D(a), \tau^C) + \Pi_m^*(t^D(a))) < 0$ . So this rises (from a positive intercept) at a decreasing rate. Either they intersect once at  $\underline{a}$ , or  $B(a) > \theta L(a)$  for all  $a$ , in which case there is never any incentive to defect, a degenerate case. ■

Note that a consequence of Lemma 2 is that if we define  $p = \Pr(C|\text{cooperation})$ ; i.e.,  $p = \Pr(B(a) < \theta L(a))$ , then  $p = \Pr(a < \underline{a})$ . Define  $\Lambda_{DS} = \frac{\delta}{1-\delta}(p^2(A - B) + p(D - N - A))$ . We now make an assumption about the likelihood that the DSP finds in favor of the defendant.

**Assumption.**  $\theta \leq \min\{\frac{\Lambda_{DS}}{L(\tilde{a})}, \frac{\Lambda_{DS}}{L(\underline{a})}\}$ .

This assumption is maintained in what follows. Notice that it is a sufficient, but not necessary, condition: if  $\frac{\Lambda_{DS}}{L(\tilde{a})}$  and  $\frac{\Lambda_{DS}}{L(\underline{a})}$  are both larger than 1, then the assumption is met by the requirement that  $\theta$  is a probability with a value bounded above by 1. Intuitively, we require there to be sufficient doubt that the DSP will find in favor of the defendant—i.e., that there is a limit to the costs associated with the use of the DSP, and therefore it is sufficiently attractive to use.

**Lemma 3.** Define  $\tilde{a}$  such that  $\theta L(\tilde{a}) = \Lambda_{DS}$ . Then  $\underline{a} \leq \tilde{a}$ .

**Proof.** From the assumption  $\theta \leq \frac{\Lambda_{DS}}{L(\tilde{a})}$ ; then  $\theta L(\underline{a}) \leq \Lambda_{DS} = \theta L(\tilde{a})$ . Now  $L' > 0$ , so  $\underline{a} \leq \tilde{a}$ . ■

**Proof of Proposition 1.** Given that foreign is playing a DSS, we must show that playing the DSS satisfies the no-defect condition for home. Given the current period draw  $\hat{a}$ , the expected current period return from defection at home is  $D(\hat{a})$ , and hence the gains from defection are  $D(\hat{a}) - \max(C(\hat{a}), D(\hat{a}) - \theta L(\hat{a})) = \min(B(\hat{a}), \theta L(\hat{a}))$ . Consider the event in which a deviation has been observed in some period. From then on, the one-shot Nash strategies are played, yielding the Nash payoff (in expectation, because the draws in the future periods are unknown) forever. That is the aggregate Nash is payoff  $V_D = \frac{1}{1-\delta}N$ . What is the foregone cooperative aggregate payoff? If cooperation occurred in the last period, in the next, each player has the option of cooperating again, or defecting. Then the value of the game in a cooperative phase is the earnings from the play in that period, plus the continuation value:

$$V = p[p(C + \delta V) + (1-p)(S + \theta L + \delta V)] \\ + (1-p)[p(D - \theta L + \delta V) + (1-p)(N + \delta V)]$$

Solving, we have  $V = \frac{1}{1-\delta}(p^2(A - B) + p(D - N - A) + N)$ . Hence,  $V - V_D = \frac{1}{1-\delta}(p^2(A - B) + p(D - N - A))$ . The no-defect condition in any period after  $\hat{a}$  is observed (and

punishment starts in the next period) is  $\min(B(\hat{a}), \theta L(\hat{a})) < \frac{\delta}{1-\delta}(p^2(A-B) + p(D-N-A))$  or  $\hat{a} < \bar{a}$ .

If  $\hat{a} < \underline{a} < \bar{a}$ , then  $B(\hat{a}) < \theta L(\hat{a})$  and the benefits of defection are too small to make either pure defection or use of the DSP mechanism worthwhile; if  $\underline{a} < \hat{a} < \bar{a}$ , the benefits of the DSP outweigh pure cooperation, but it is still intertemporally optimal to voluntarily pay the proportionality penalty to benefit from the possibility of cooperation in the next period. The no-defect condition is violated when  $\hat{a} > \bar{a}$ ; then the gains from pure defection, and the Nash reversion play from then on are preferred to cooperation. Hence, a pair of DSSs is an equilibrium. ■

**Proof of Proposition 2.** The assumption implies  $\theta L(\bar{a}) \leq \Delta_{DS} = \theta L(\bar{a})$ . Since  $L' > 0$ ,  $\bar{a} < \bar{a}$ . ■

**Proof of Corollary 1.**  $\frac{\theta L(a^{\max})}{\Delta_{DS} + \theta L(a^{\max})} < \frac{B(a^{\max})}{\Delta_{PD} + B(a^{\max})} \iff \theta < \frac{\Delta_{DS}}{L(a^{\max})} \cdot \frac{B(a^{\max})}{\Delta_{PD}}$ . Now  $\theta \leq \frac{\Delta_{DS}}{L(\bar{a})}$ , and  $L(\bar{a}) < L(a^{\max})$ , so  $\theta < \frac{\Delta_{DS}}{L(a^{\max})}$ . Clearly,  $\frac{B(a^{\max})}{\Delta_{PD}} > 1$ . The set of discount factors under which the standard PD under uncertainty can support a cooperative equilibrium is  $(0, \frac{B(a^{\max})}{\Delta_{PD} + B(a^{\max})}) \subset (0, \frac{\theta L(a^{\max})}{\Delta_{DS} + \theta L(a^{\max})})$ , the set of discount factors for which a DSS equilibrium exists. ■

## REFERENCES

- Bagwell, Kyle, and Robert W. Stager. 1999. "An Economic Theory of the GATT" *American Economic Review* 89 (March): 215–48.
- Baldwin, Richard. 1987. "Politically Realistic Objective Functions and Trade Policy." *Economics Letters* 24 (3): 287–90.
- Bello, Judith Hippler. 1996. "The WTO Dispute Settlement Understanding: Less is More." *American Journal of International Law* 90: 416–18.
- Bhagwati, Jagdish and Hugh T. Patrick (eds.) 1990 *Aggressive Unilateralism: America's 301 Trade Policy and the World Trading System*. Ann Arbor, MI: University of Michigan Press.
- Bown, Chad P. 2004. "On the Economic Success of GATT/WTO Dispute Settlement." *The Review of Economics and Statistics* 86 (August): 811–23.
- Busch, Marc. 1999. "Forum Shopping for Dispute Resolution in US–Canada Trade." Presented at International Studies Association Meetings, Washington DC.
- Chayes, Abram, and Antonia Handler Chayes. 1993. "On Compliance" *International Organization* 47 (Spring): 175–205.
- Downs, George W., and David M. Rocke. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton, NJ: Princeton University Press.
- Downs, George W., D.M. Rocke, and P.B. Barsoom. 1996. "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50 (Summer): 379–406.
- Dunoff, Jeffrey L., and Joel P. Trachtman. 1999. "Economic Analysis of International Law." *Yale Journal of International Law* 24 (1): 1–59.
- Ethier, Wilfred J. 2001. "Punishments and Dispute Settlement in Trade Agreements" PIER Working Paper No. 01-021.
- Fearon, James D. 1998. "Bargaining, Enforcement and International Cooperation." *International Organization* 52 (Spring): 269–305.
- Garrett, Geoffrey, and James McCall Smith. 2002. "The Politics of WTO Dispute Settlement." UCLA International Institute. Occasional Paper Series.
- Goldstein, Judith, and Lisa L. Martin. 2000. "Legalization, Trade Liberalization, and Domestic Politics: A Cautionary Note" *International Organization* 54 (Summer): 603–32.
- Goldstein, Judith, Miles Kahler, Robert Keohane, and Anne-Marie Slaughter. 2000. "Introduction: Legalization and World Politics." *International Organization* 54 (Summer): 385–400.
- Grossman, Gene M., and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review* 84 (September): 833–50.
- Gruber, Lloyd G. 1999. "Rethinking the Rational Foundations of Supranational Governance: Lessons from the North American Free Trade Agreement." Presented at International Studies Association Meetings, Washington, DC.
- Hoekman, Bernard M., and Michel M. Kostecki. 2001. *The Political Economy of the World Trading System*. 2nd ed. New York: Oxford University Press.
- Jackson, John H. 1997a. "Editorial Comment: The WTO Dispute Settlement Understanding—Misunderstandings on the Nature of Legal Obligation." *American Journal of International Law* 91: 59–64.
- Jackson, John H. 1997b. *The World Trading System*. 2nd ed. Cambridge: MIT Press.
- Jackson, John H. 1998. "Designing and Implementing Effective Dispute Settlement Procedures. WTO Dispute Settlement, Appraisal and Prospects." In *The WTO as an International Institution*, ed. Anne O. Krueger. Chicago: Chicago University Press, chap. 5, 161–80.
- Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.
- Kovenock, Dan, and Marie Thursby. 1994. "GATT, Dispute Settlement, and Cooperation." In *Analytical and Negotiating Issues in the Global Trading System*, ed. Alan V. Deardorff and Robert M. Stern. Ann Arbor: Michigan University Press, 361–90.
- Lawrence, Robert Z. 2003. *Crimes and Punishments: Retaliation Under the WTO*. Washington, DC: Institute for International Economics.
- Levy, Philip I., and T. N. Srinivasan. 1996. "Regionalism and the (Dis)Advantage of Dispute Settlement Access." *American Economic Review* 86 (May): 93–8.
- Maggi, Giovanni. 1999. "The Role of Multilateral Institutions in International Trade Cooperation." *American Economic Review* 89 (March): 190–214.
- Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2000. "Free to Trade: Democracies, Autocracies and International Trade." *American Political Science Review*, 94 (June): 305–322.
- Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2002. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Organization* 56 (Fall): 477–514.
- Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics and Politics* 2 (March): 1–23.
- Milner, Helen V., and B. Peter Rosendorff. 1996. "Trade Negotiations, Information and Domestic Politics." *Economics and Politics* 8 (July): 145–89.
- North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press.
- Oye, Kenneth. 1986. *Cooperation Under Anarchy*. Princeton, NJ: Princeton University Press.
- Ozden, Caglar. 2001. "Trade Agreements Under Asymmetric Information and Efficiency Gains from WTO's Dispute Settlement Process." Manuscript, Emory University.
- Pevehouse, Jon C., Emile Hafner-Burton, and Matthew Zierler. 2002. "Regional Trade and Institutional Design: Long after Hegemony?" Presented at the 2002 MPSA Meetings, Chicago, IL.
- Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42 (Summer): 427–60.
- Reinhardt, Eric. 1999. "Aggressive Multilateralism: The Determinants of GATT/WTO Dispute Initiation, 1948–1998." Presented at International Studies Association Meetings, Washington.
- Reinhardt, Eric. 2001. "Adjudication Without Enforcement in GATT Disputes." *Journal of Conflict Resolution* 45 (April): 174–95.
- Rosendorff, B. Peter, and Helen V. Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55 (Fall): 829–57.
- Setear, John K. 1997. "Responses to Breach of a Treaty and Rationalist International Relations Theory: The Rules of

- Release and Remediation in the Law of Treaties and the Law of State Responsibility." *Virginia Law Review* 83: 1–127.
- Smith, James McCall. 2000. "The Politics of Dispute Settlement Design. Explaining Legalism in Regional Trade Pacts" *International Organization* 54 (Winter): 137–80.
- Staiger, Robert, and Guido Tabellini. 1999. "Do GATT Rules Help Governments Make Domestic Commitments?" *Economics and Politics* 11 (July): 109–44.
- Sykes, Alan. 1991. "Protectionism as a Safeguard" *University of Chicago Law Review* 58: 255–305.
- Trebilcock, Michael J., and Robert Howse. 1999. *The Regulation of International Trade*. 2nd ed. New York: Routledge.
- Yarbrough, Beth V., and Robert M. Yarbrough. 1997. Dispute Settlement in International Trade: Regionalism and Procedural Coordination. In *The Political Economy of Regionalism*, ed. Edward D. Mansfield and Helen V. Milner. New York: Columbia University Press, 134–63.

# Reading Habermas in Anarchy: Multilateral Diplomacy and Global Public Spheres

JENNIFER MITZEN *Ohio State University*

*States routinely justify their policies in interstate forums, and this reason-giving seems to serve a legitimating function. But how could this be? For Habermas and other global public sphere theorists, the exchange of reasons oriented toward understanding—communicative action—is central to public sphere governance, where political power is held accountable to those affected. But most global public sphere theory considers communicative action only among nonstate actors. Indeed, anarchy is a hard case for public spheres. The normative potential of communicative action rests on its instability: only where consensus can be undone by better reasons, through argument, can we say speakers are holding one another accountable to reason. But argument means disagreement, and especially in anarchy disagreement can mean violence. Domestically, the state backstops argument to prevent violence. Internationally, I propose that international society and publicity function similarly. Public talk can mitigate the security dilemma and enable interstate communicative action. Viewing multilateral diplomacy as a legitimation process makes sense of the intuition that interstate talk matters, while tempering a potentially aggressive cosmopolitanism.*

How could multilateral diplomacy—"talk" and argument among states—legitimate state action? Scholars, practitioners, and the broader public commonly link an international action's legitimacy to the multilateral diplomacy that surrounded it. NATO's intervention in Kosovo is widely perceived as legitimate, in large part because of the arguments advanced in the diplomacy before and immediately after (Johnstone 2004); conversely, the American-led coalition's Iraq War is widely perceived as illegitimate in part because of the way the United States conducted its multilateral diplomacy (Rubin 2003). In short, we take for granted that public, interstate talk matters for legitimacy; it is part of our common sense about contemporary world politics.

The problem is that it is not clear how talk could matter for legitimation in an anarchic system, because argument is an inherently unstable social practice. As developed especially by Jürgen Habermas (1984, 1996), argument may be defined as the exchange of reasons by participants who are oriented to reaching consensus and remain open to changing their minds if faced with better reasons. Habermas links argument normatively to communicative action, the promise of which is that consensus resulting from argument will be for the right reasons, i.e., reasons that are good for the collective and not simply for the most powerful. Importantly, however, this normative potential rests on a fundamental instability. Since it only is possible to say speakers are

holding one another accountable to reason if agreements can be undone through future argument, each consensus must remain contingent. That is, in its ideal form, argument has the power to undo and remake social consensus. Yet this power can by its nature lead to violence. In fact, argument is usually a precursor to violence, suggesting that argumentative processes face a potentially slippery slope. Without some constraint to keep actors committed to resolving their disagreements discursively, argument can spill over from the conference table to the street, or even to the battlefield. To sustain argumentative legitimation, then, an environment must be able to contain the instability of communicative action. It must permit argument while guarding against the potential that argument will degenerate into violence.

This makes anarchy a hard case for the proposition that public talk could legitimate state action. Habermas conceptualizes communicative action in the context of a consolidated democratic state, which blocks the slippery slope to violence. But anarchy lacks any such centralized prevention of violence, and as such the slippery slope from argument to violence is very much in force. Unlike the domestic case, in anarchy there is no easy answer to Bent Flyvbjerg's (1998, 80) question, "Why use the force of the better argument when force alone will suffice?" In short, if in anarchy argument can easily descend into violence, how could the multilateral diplomacy surrounding the Iraq War, or any interstate talk, be anything other than cheap, a rhetorical veneer to interests and power, incapable of "legitimizing" anything?

Despite the problems with argument in anarchy, in recent years a substantial literature has emerged in both normative and explanatory theory on argument and communicative action in world politics. Normative theorists have become interested in global public spheres, discursive structures that enable communicative action beyond state borders. Global public spheres hold out the prospect that democratic self-governance, governance that aims for the collective good, could

Jennifer Mitzen, Assistant Professor, Ohio State University, 2140 Derby Hall, 154 North Oval Mall, Columbus, OH 43210-1373, mitzen.1@osu.edu

For comments on previous drafts, many thanks go to James Bohman, Cosette Creamer, George Gavriliis, Yoram Haftel, Patrick Jackson, Hans-Martin Jaeger, Zack Kertcher, Andrew Linklater, Patchen Markell, Anthony McGrew, Amanda Metskas, Lee Sigelman, Lisa Wedeen, Nick Wheeler, Iris Young, and especially Michael Neblo and Alexander Wendt. Earlier versions were presented at the University of Chicago's Globalization Workshop, the International Studies Association Conference, New Orleans, Louisiana, March 2002, and the British International Studies Association Conference, London, England, December 2002.

extend to a scale beyond anything humankind has known before. Importantly, however, in this work legitimation is assumed to take place only among citizens and directed toward states, rather than among states themselves. It is a vertical, not a horizontal, process. The interstate dimension plays at most an ordering role, not a legitimating one (e.g., Bohman 1999; Habermas 1997).<sup>1</sup> In contrast, building especially on the work of Thomas Risse (2000), empirical theorists of international relations (IR) (e.g., Müller 2001; Payne 2001) and international law (Brunée and Toope 2000; Johnstone 2004), have begun to examine interstate argument as a legitimation process. But this literature on the would-be horizontal dimension of global public spheres has not come to grips with the difficulty of containing argument in anarchy. The institutional prerequisites for horizontal public spheres have not been theorized, and so it is difficult in the end to fully accept that public, interstate argument could legitimate state action.

In my view there are two reasons to consider multilateral diplomacy as a dimension of global public sphere legitimation. Empirically, excluding interstate talk does not make sense of our intuitions that multilateral diplomacy “matters.” If talk in IR is always cheap, then it is not clear why states would bother to talk at all. Normatively, excluding multilateral diplomacy strips it and the institutions of the states system that enable it of value. If talk in IR cannot legitimate, then it is not clear why states *should* bother to talk. Moreover, as we shall see, excluding multilateral diplomacy permits, if not encourages, a potentially aggressive cosmopolitanism.

With these stakes in mind, I confront the slippery slope of anarchy to conceptualize multilateral diplomacy as the horizontal dimension of global public spheres. I propose that, in the contemporary international system, the instability of communicative action is contained by what I call the “forum effects of talk,” which thereby make horizontal argumentative legitimation possible. The forum effects are sustained by the institutions of international society and by publicity, where publicity refers to both opportunities for face-to-face engagement and a more mediated visibility made possible by communications technologies. Extant global public sphere theory tends to focus on the latter, which has expanded the nonstate audience of state behavior. I focus instead on face-to-face visibility among states, which in the form of conference diplomacy was introduced to the system in the early nineteenth century when the European Great Powers decided to jointly manage the balance of power. The upshot is that global public spheres have two legitimation dynamics: a widely recognized vertical one centered on the practices of cosmopolitan citizens and transnational nonstate actors, and a neglected horizontal one among states, which I call an “interstate” pub-

lic sphere. I illustrate the theoretical claims with an example from the Concert of Europe, a case which, while by no means a fully realized public sphere, is not as strange for public sphere theory as it might seem. In conclusion I consider the implications of viewing multilateral diplomacy as a legitimation process by suggesting how it might affect an analysis of the diplomacy surrounding the Iraq War.

## THE GLOBAL GOVERNANCE TWO-STEP

Confronting the limitations of nation-state democracy in conditions of globalization, Habermas (1998, 2001) and other theorists argue that it is necessary today to think in terms of state-transcending—global or cosmopolitan—rather than just national public spheres.<sup>2</sup> This new thinking faces the challenge of containing communicative instability, or maintaining a context that can permit new and better arguments to emerge without constantly threatening to descend into violence. In Habermas’ domestic theory, the state’s centralized power plays this crucial role by protecting physical safety. Of course, it matters normatively whether the state is democratic or authoritarian, and some states have no public spheres whatsoever. But the existence of a state—a solution to the Hobbesian problem—is never in question. Argumentative practices might democratize an authoritarian state (Habermas 1994b), but they never threaten to throw the population back to the state of nature. This suggests what could be called “two-step” reasoning about global public spheres (cf. Legro 1996). All governance requires both social order and legitimation. When theorized as a two-step, order is considered to be supplied separately or exogenously from legitimation. This is not necessarily bad. Indeed for domestic public sphere theory it may make sense to bracket state consolidation and assume social order. But in anarchy we cannot so easily take order for granted. The instability of communicative action thus poses more of a problem for global public spheres, which suggests that two-step reasoning would be counterproductive at this level. In this section, focusing especially on Habermas, I first show how the state contains the instability of communicative action in the domestic context, and then examine extant strategies for stabilizing communication in the international context. In each strategy, the production of order precedes and remains separate from legitimation. Even where global public spheres rest on interstate cooperation, only arguments by nonstate actors can legitimate state power. Interstate dynamics are associated, at best, with the production of order.

<sup>1</sup> I focus on Habermasian theory, but there also is a Deweyian strand of global public sphere theory (e.g., Cochran 2002), some of which incorporates interstate processes (e.g., Brunkhorst 2002). Still, the majority of even that work stresses vertical legitimation.

<sup>2</sup> Global public sphere terminology varies. To simplify, I use “global” to refer to all state-transcending public spheres and propose that global public spheres are characterized by two levels: “transnational” public spheres, constituted by vertical, critical dynamics among nonstate actors, and “international” public spheres constituted by horizontal dynamics among states.

## Communicative Action In Domestic Public Spheres

Communicative action, or the exchange of reasons oriented toward understanding, is the heart of public sphere theory. Communicative action builds from the premise that reason is intersubjectively constituted and inheres in linguistic communication. In everyday utterances, speakers raise validity claims—claims about what is objectively true or morally right for the group—and there is a tacit, shared expectation that, if challenged, a speaker can offer acceptable reasons. The exchange of validity claims constitutes the process of argument, and consensus resulting from such argument is the ideal form of social integration. Habermas develops this idea to counter the pessimistic, Weberian narrative of modernity as the triumph of strategic action and instrumental, technical rationality. For Habermas, modernity also has given rise to a new emancipatory potential for self-governance based on reason in public spheres.

Communicative action embodies an inherent tension between social acceptance, or stability, and validity. On the one hand, it can generate stable consensus: a positive response to a validity claim creates an agreement on a fact and “obligations relevant to future interaction” (Habermas 1996, 20). But validity claims also always point beyond a particular context. Because ideally, communicative agreements are supported by the “best” reasons, any achieved agreement must remain open to “better” reasons in the future. Through social learning and change, over time some reasons can become obsolete. Only where argument can undo previous agreements is it possible to say speakers are holding one another accountable to reason.

In public sphere theory, the public’s communicative action, which Habermas calls public reason, can hold material or social power accountable. In so doing, public reason is not just a “check” on material power but changes its nature. Insofar as political power justifies its use according to public reason, therefore, one can say that political power has been drawn out from its material locale and lodged in the communicative power of those affected. This is the emancipatory promise of public sphere theory. To attain that promise, public spheres in practice have two dimensions or “tracks” (Habermas 1996, chap. 8). The informal or “critical” public sphere is characterized by a vertical dynamic of subjects holding decision makers accountable; Habermas calls it a “transmission belt” of social concerns to decision-making bodies. The formal or decision-making sphere, in contrast, is characterized by a horizontal dynamic; it exists once a state has a parliament or congress, which infuses the decision-making process itself with reason giving and justification (Fraser 1992).<sup>3</sup> In a functioning public sphere, public reason is salient both outside formal decision-making bodies and within them.

Governance through public reason is demanding. First, speakers must recognize one another’s communicative competence and grant each other the right to disagree. Second, they must approach interaction with an orientation to listen—to reflect on others’ arguments rather than simply coerce them or engage in violence. That is, they must commit to the process of argument, which means they will not let the fact of disagreement destroy the group (Habermas 1984, 36–37; Habermas 1996, 20–21; White 1994, 35 ff.). With these in mind, it is easy to see that public sphere governance places strict demands on the social environment where argument takes place. Perhaps the most minimal condition is that speakers feel confident of their physical safety. Where individuals face the constant risk of violence they cannot reflect or listen, much less argue; all energy is consumed with securing survival (Mitzen n.d.). The public sphere environment must therefore encourage the orientation to listen, which means that, while permitting disagreement, it must also somehow contain it, preventing disagreement from spilling over into violence.

Habermas argues that the best environment for public spheres is a vast reserve of shared background knowledge. He calls the sphere of interaction organized around such consensual knowledge a lifeworld, and includes institutions such as religion and the family, moral norms, and cultural practices. This “culturally familiar,” unproblematic environment helps “explain how the daily process of consensus building is time and again able to cross the threshold of the risk of dissent” (cited in Müller 2001, 169). A shared normative context provides safety by giving decision makers the motivation for self-restraint and citizens the motivation to participate rather than withdraw or rebel. Historically, lifeworld contexts were so fully internalized by members as to be rarely reflected on. But modernity has rationalized them, i.e., differentiated social life in such a way that aspects of the tacit background consensus can be brought into public light and debated. This means they can become subject to collective reason and the force of better arguments. Rationalization makes public spheres possible by maintaining the safety of the lifeworld while injecting new potential for reflection and argument.

The problem is that rationalized lifeworlds do not form a thick basis for modern political groups. Instead, modern groups are characterized primarily by complexity and pluralism. Markets, bureaucracies, and powerful political systems embed individuals in a complex web of relations they are not fully aware of and cannot extricate themselves from. In one sense this behind-the-back integration is useful: markets and bureaucracies maintain social cohesion in contexts where members would otherwise be “overburdened in their efforts at reaching understanding” (Habermas 1996, 38). But markets and bureaucracies tend to expand and, unlike lifeworlds, neither requires a communicative consensus to function. As these systems come to structure more of social life, Habermas argues that they can squeeze out the potential for public reason. In addition, members of modern political groups do not

<sup>3</sup> Fraser (1992) uses “weak” and “strong” to refer to critical and decision-making public spheres



generally share thick lifeworld bonds; they are “strangers.” Strangers might not see consensus as desirable; they might not recognize one another as capable of communicative consensus at all, much less be willing to listen and reflect on each other’s arguments. Among strangers, the potential for violence is harder to contain. Indeed, it can be so close to the surface that argument becomes impossible. Even if an individual wants to listen, it is impossible to know the other’s intentions, and so each claim raised by each side in argument raises the specter of violence.

In the modern context, then, sustaining the potential for public reason is a serious challenge. Habermas offers two responses. First, he argues that political deliberation should be modeled as a mix of moral, ethical, and pragmatic discourses (1996, 165–7). In political deliberation, decision makers first determine whether to address a given question through moral or ethical argument, or through bargaining. The latter is valid where participants determine that no general interest or shared value is at stake. Bargaining is of course not communicative action: here, social power is not “neutralized” but manifest in threats and promises. Still, bargaining can be “fair,” and this criterion maintains the link between political deliberation and moral discourse. If bargaining procedures are deliberative and justifiable in moral discourse, as long as their outcomes are contingent, then “understanding beyond instrumental-rational agreement is possible” (2001, 109). In short, fair bargaining is a normative achievement, implying deliberation that is constrained by norms. Fair bargaining means that, even in conditions of complexity and pluralism, political deliberation maintain the normative potential of communicative action.

Second, Habermas anchors public spheres in law, by which he means positive law, law that is legislated and enforced. Law is unique in its capacity to convert normative ideals to social facts. This is because, for one, legal rules share the instability of communicative action by maintaining the potential for a gap between the socially accepted rules and the best rules. Although invoked in a particular context, legal rules always point beyond that context to a larger, general interest that can be rendered in moral terms as the common good (Bohman 1994, 899; Habermas 1984, 81, 178). But positive law contains this instability, lessening the potential that argument will spill over into violence, because law is centrally enforced and as such exists irrespective of whether citizens legitimate it in a particular instance. The legal system therefore can be seen as a safety net for communicative action. Enforcement allows “convictions to be replaced by sanctions in that it leaves the motives for rule compliance open while enforcing obedience” (Habermas 1996, 38, 448–9).

Habermas’ analysis of law is complex and nuanced, and certainly he is not arguing that its enforcement (facticity) trumps or opposes its legitimacy (validity). Indeed, he argues that law can only anchor public spheres if it is democratically generated and if its contents protect the preconditions for communicative action such as the rights to privacy, equality, and partici-

pation. Thus, not all states can sustain public spheres. Still, the state’s centralized enforcement is essential to the logic. Although the stability law provides differs from that of the lifeworld—its link to communicative action is “artificially” produced by sanctions rather than “organic(ally)” produced by “inherited forms of life” (1996, 30)—Habermas argues that its capacity to secure communicative action and the capacity to compel compliance are as internally or logically related as they are in lifeworld contexts. “A force that otherwise stands opposed to the socially integrating force of communication (i.e., centralized coercion) is, in the form of legitimate coercion, thus converted into the means of social integration itself” (1996, 462). In short, the state’s enforcement power is crucial to making public reason possible in modern political life.<sup>4</sup>

This emphasis on centralized enforcement might just be an artifact of the domestic origins of public sphere theory. In Habermas’ ([1962] 1994b) historical narrative, public spheres emerged within existing European states with the express purpose of democratizing them. It is then no surprise that Habermas’ template for public spheres assumes a context in which enforcement is possible. Still, the role of state enforcement in Habermas’ account is significant for two reasons. First, it analytically separates the production of order from the production of legitimacy, making domestic public sphere theory reliant on two-step reasoning. Public spheres require an already-existing centralized power. The theory brackets how that enforcement capacity is formed and reproduced and instead studies its role in making legitimation possible. Second, Habermas contrasts enforced modern law to customary premodern law and does not consider the possibility of a law in modernity that lacks centralized enforcement. This has important consequences for how he theorizes global public spheres, in effect ruling out a priori the possibility that international law might stabilize social life in an analogous way to enforced law.

### Communicative Action in Global Public Spheres

The challenge for global public spheres is how to contain the instability of communicative action where argument is not backstopped by either a shared lifeworld or positive law. Extrapolating directly from the domestic context, global public spheres would require world government: a supersovereign power capable of enforcing cosmopolitan law. Although at times Habermas’ writings in the 1990s suggest this as a distant but hopeful possibility, two other strategies for maintaining order figure more prominently in his work and that of other global public sphere theorists: the democratic peace and international regimes (including international organizations). These three strategies are not mutually exclusive. Outlining how each operates in Habermas’ theory, it becomes clear that, even as the

<sup>4</sup> The argument is not without criticism, some commentators question whether this template sacrifices the radical potential of public sphere theory, e.g., Bohman (1994)

positivity condition is relaxed, two-step reasoning persists. Because it lacks enforcement, the anarchic world of states is seen as a dangerous balance of power where the most we can hope for is order, not legitimation. Multilateral diplomacy, if factored in at all, is consigned to the realm of order production and plays no role in legitimation, which is theorized in purely vertical terms.

**Global Positive Law.** One strategy for global public spheres suggested by Habermas is to expand positive law at the international level. The premise is that, globally as much as domestically, the protection of individual human rights is necessary to contain communicative instability. Because the state is often the culprit in human rights violations, citizens need to be able to make claims against their states, which global positive law can help ensure. As in the domestic case, enforcement is crucial. Therefore, a global "executive power" is needed to intervene authoritatively where human rights violations have occurred. "The community of peoples must at least be able to hold its members to legally appropriate behavior through the threat of sanctions. Only then will the unstable system of states asserting their sovereignty through mutual threat be transformed into a federation whose common institutions take over state functions: it will legally replace the relations among its members and monitor their compliance with its rules" (1997, 127).

Habermas recognizes that a world of enforced individual rights is a long way off. The contemporary United Nations' (UN's) hybrid status as an institution premised on both sovereignty and human rights means that currently it can have only the minimal agenda of preventing war and reacting to human rights abuses (2001, 107–8). Still, the organization can be strengthened. He calls for a stronger UN with a military force to implement decisions and UN reform to expand the role of the Security Council and strengthen the International Criminal Court (1999a: 268). An improved UN would serve as one leg of a system of multilevel governance analogous to the European Union (1998). If foreign policy is the realm of unregulated violence, and domestic policy the realm of rights and regulation, then the new era would be one of "multilaterally coordinated world domestic policy" (1994a, 23–4).

Two aspects of this argument stand out. First, if global public spheres ultimately require global positive law, then sovereignty and the states system would seem to be problems to be overcome in global governance rather than essential to its legitimation. Indeed, Habermas' proposals to strengthen the UN would effectively end state sovereignty, with increased centralization of military/executive, legislative, and judicial powers at the global level. Second, the argument extrapolates directly from the domestic template, which means that like domestic public sphere theory it brackets how enforcement power is consolidated. Enforcement capacity might expand as a result of deliberative processes, as in UN reform. But it could also happen through imposition or force, and the use of military force to expand the sphere of enforced rights can be hard for others to distinguish from liberal imperialism.

**The Democratic Peace.** A second strategy to anchor global public spheres, also found in Habermas and echoed by other theorists (e.g., Bohman 1999; Erikson and Fossum 2000), is through the spread of liberal democracy at the national level. In one sense this is an aggregative logic: public spheres are a democratic ideal, and so spreading democracy expands public spheres. As democracy spreads, citizen-based associations and nongovernmental organization achieve greater roles and reach across boundaries, giving cosmopolitan values increasing prominence (Bohman 1997, 196–7; Habermas 1997, 125). Such groups and linkages can grow only where citizens have political voice and freedom of association, rights that are associated with liberal democracy. Liberal political culture is the "ground in which the institutions of freedom put down their roots" and "medium" to achieve that progress, and can only be forged globally through the proliferation of democratic states (Habermas 1997, 125; Habermas 2001, 111–12). Moreover, the proliferation of states that enforce democratic rights at the domestic level translates to less need for global enforcement.

But sovereignty complicates any simple aggregation of national into global public spheres, because, irrespective of a state's regime type, as a sovereign state it must survive in the competitive, potentially dangerous environment of anarchy. As such, expanding the number of democratic states can only create global public spheres if the competitive dynamics among states can be dampened. With this in mind, global public sphere theory invokes the democratic peace, the finding in IR scholarship that democracies tend not to fight one another (e.g., Doyle 1986). Because democracies can be counted on not to fight, they form a "zone of peace," the semblance of a transnational community (e.g., Bohman 1997, 180–1; Habermas 1998). Indeed, this work tends to associate the spread of democracy with deeper, more durable interstate cooperation in all issue areas (see also Slaughter 1995). At the same time, between democracies and nondemocracies there remains a balance-of-power world where peaceful intentions of others cannot be assumed (Habermas 1997, 131–2).

Importantly, Habermas interprets the democratic peace as rooted in purely internal or domestic dynamics: cosmopolitan citizens of liberal democracies cannot be mobilized for war against fellow democracies (1997, 120–1). The peace is therefore induced vertically, by civil societies holding their decision makers accountable, which happens as publics of individual states incorporate "higher order value orientations" into their preferences and press leaders to pursue those values (1999b, 451–2). Moreover, like the peace, deep cooperation more generally among democracies also has unit level roots. Democracies cooperate well internationally because each individually is committed to the rule of law and tends to comply with agreements. When both parties to an agreement have a domestic political culture encouraging compliance, compliance is more likely. This means the democratic peace, and interdemocratic cooperation, need not be consciously constructed or sustained by international institutions.

It certainly is true that the spread of democratic regimes would strengthen already-existing public spheres, insofar as it would guarantee conditions of communication for more individuals than currently are able to participate. But because of anarchy and sovereignty it is not clear that spreading democracy could create, or that the existence of democracies alone could sustain, conditions for global communicative action. In fact, three aspects of the democratic peace suggest that it ought not be taken as a necessary precondition for global public spheres.

First, it is not clear that the democratic peace is necessary to solve the problem of war. For example, it is not the case that liberal states maintain a zone of peace while nonliberal states inhabit a realist world. Stable interstate peace has evolved among states of various regime types, starting in the nineteenth-century Concert of Europe, and interstate war has declined systemwide since 1945. Second, as José Alvarez (2001, 200 ff.) argues, the zone-of-peace argument implies greater compliance with international law among democratic states than in cooperation among states of mixed regime types. But the empirical record shows that regime type matters little for compliance: liberal states are not necessarily more law-abiding. Finally, relying on a specifically unit-level explanation of the democratic peace to anchor global public spheres maintains two-step reasoning, where order is produced separately from legitimation. Then, because the order-production logic is independent of the states system, it is easy to focus solely on bottom-up processes and delegitimate horizontal practices and norms such as multilateral diplomacy.

Importantly, whereas Habermas relies on a unit-level causal explanation for the democratic peace, there is in fact a debate in IR about whether the democratic peace is rooted in unit- or system-level dynamics. Other versions of the democratic peace stress systemic factors rather than just internal ones (e.g., Cederman 2001). Moreover, it certainly is plausible that international law itself helps cause the democratic peace, because it is not clear if democracies would behave peacefully toward each other in a world without it. Note that this is not a question of whether we can trust the democratic peace as a real empirical phenomenon, an issue about which there also is a great deal of controversy (e.g., Rosato 2003). The issue for global public sphere theory is, granting the democratic peace, how should we explain it? A unit-level explanation treats international law and multilateral diplomacy as irrelevant whereas a system-level explanation does not.

In other words, as long as sovereignty remains, the choice to ground global public spheres in a unit-level logic that depends on state regime type has the implication of excluding nondemocratic states from global governance. This does little to ground communicative action between democratic and nondemocratic states and is vulnerable to the suggestion that forceful intervention by existing democracies to create democratic regimes is always justified.

**Interstate Regimes.** Habermas' third strategy to contain the instability of communicative action in anarchy, and perhaps the most popular among other theorists (e.g., Bohman 1999; Linklater 1998; Lynch 1999), examines public sphere formation in the context of international institutions or "regimes" (Krasner 1983), where participating states ideally are, but may not be, democracies. The strategy builds on the neo-utilitarian logic of rationalist IR regime theory, which shows how cooperation can emerge in anarchy among self-interested states (e.g., Keohane 1984; Ruggie 1998). But rather than focus on how institutions mitigate the security dilemma, the focus of IR literature, Habermas and others stress that, by broadening the audience of state behavior, these interstate institutions can become locales for the transnational exchange of reasons and opinion formation. As Bohman (1999: 500) puts it, regimes provide a "practical foothold" and potential infrastructure for cosmopolitan democracy. Insofar as they bring nongovernmental organizations into interstate bargaining processes, for example, citizens can increasingly hold states accountable for actions on the international as much as the domestic stage.

But whereas this work acknowledges an important role for states in global public spheres in making cosmopolitan democracy possible, it retains two-step reasoning. States themselves provide only order, not legitimation. Indeed, for Habermas, international regimes are barely one step removed from the power politics of a Hobbesian state of nature. In his words, state decisions in organizations such as the World Trade Organization and World Bank are no more than "naked" compromise formation that simply reflects back the essential features of classical power politics; such communication cannot reflect or develop any 'thick' communicative embeddedness" (2001, 109). Even Bohman, who builds more explicitly from IR's regime theory, essentially comes to the same conclusion that interstate decision making is not linked to communicative action. Thus, whereas domestic public spheres have two dimensions, vertical and horizontal, global public spheres are characterized only by a vertical dimension.

Interestingly, this strategy for anchoring global public spheres renders political bargaining and compromise among states fundamentally different than at the domestic level. As we have seen, at the domestic level Habermas accepts that political deliberation is characterized as much by bargaining and compromise as it is by argument and moral discourse, but argues that "fair bargaining" maintains a connection to communicative action and as such is part of public spheres. He does not, however, extend this reasoning to interstate bargaining. Habermas acknowledges that "normative framing conditions" might shape a state's "choice of rhetoric" and help structure international negotiations, but the origin of those framing conditions is not clear. Indeed, since he stresses the close link between interstate talk and balance-of-power politics, it would seem that any norms states follow rhetorically would have to be derived from and aimed at domestic audiences alone. For Habermas (2001, 71), communicative action cannot take place among states, particularly where states do

not share regime type. International agreements simply cannot have legitimating force and can never rise above compromise.

Building global public spheres from international regimes is not itself problematic; indeed my account of global public spheres similarly begins by examining efforts at interstate cooperation. The error is to exclude such cooperation from public sphere dynamics and conclude that global legitimation processes are purely vertical. The problems with that conclusion mirror those with the democratic peace.

First, it maintains two-step reasoning, attributing two separate logics to the production of order and legitimation at the global level. But note that, again, this two-step rests on a particular, in this case rationalist, interpretation of cooperation's causes and dynamics. In fact, as with the democratic peace, there is a debate in IR about how to best understand international regimes. Rather than adopt utilitarian logics that maintain broadly realist assumptions, many IR scholars assume that states interact in a normatively much thicker environment—an international society, culture, or even a community (e.g., Kratochwil 1989; Wendt 1999). This constructivist approach to regimes sees the day-to-day rhetorical practices among states in regimes as largely communicative, which suggests the possibility of horizontal legitimation, whereas a utilitarian approach does not. Second, excluding interstate linguistic processes from global public spheres scales back their emancipatory potential, because, unlike parliaments in the domestic case, here global decision making itself does not get democratized. If globally there are at most critical publics emerging from civil societies, public reason has at most a reactive, countersteering role.<sup>5</sup>

Why is it so difficult for Habermasian global public sphere theorists to see multilateral diplomacy as a way to legitimate state action? The answer might be normative: they might feel sovereignty is outmoded and ought not anchor global governance. It might be inadvertent: they may be simply transposing the existing public sphere template onto the international environment without thinking about the distinctive problems of anarchy. Or it might be philosophical: they might object to the notion that a corporate actor like the state could engage even in principle in communicative action (see Wendt 2004). It is hard to say, because none of this work treats the issue explicitly. It is simply assumed that states cannot engage in communicative action. Into this silence, I have offered a principled reason for the exclusion of states from public sphere theory, rooted in public sphere theory itself, which unifies the literature and suggests a pathway toward a solution. Namely, the need to contain communicative instability leads theorists to two-step reasoning. Where order is produced in a different sphere than legitimation, legitimation can fail without necessary repercussions for social order: the ability to keep the conversation going is never in

question. But, in fact, it is hard to keep order and legitimation so distinct. Every social order is intimately tied to legitimation processes, because durable order always rests on a consensus regarding the truth of particular value claims. Because authoritative decisions implicate these values, legitimation processes always either support or undermine order. This suggests the need to look beyond the two-step for other ways to contain communicative instability in global governance. That search, in turn, leads back to the states system.

## INTERNATIONAL PUBLIC SPHERES

Unlike global public sphere theorists, a number of IR scholars have been viewing regimes as sites for communicative action (e.g., Ellis 2002; Samhat and Payne 2003) and persuasion (e.g., Checkel 2001; Johnston 2001). A few have explicitly explored the preconditions for communicative action among states (e.g., Müller 2001; Risse 2000) and even the possibility of Habermasian discourse ethics on a global scale that includes dialogue among states (e.g., Linklater 1998). None, however, has directly confronted the instability of communicative action and the problem of violence it raises, nor has this work moved beyond two-step reasoning to establish structural conditions for international public spheres. Indeed, the difficulty of containing communicative action in anarchy is a strong theoretical challenge to this literature. Anarchy is a harder case for communicative action than the literature has acknowledged. Even where states want cooperation, it is hard to secure. A major impediment is mistrust at a structural level: the security dilemma. States cannot be sure of one another's intentions, and they draw on the same repertoire of actions to defend themselves as they do to aggress. The security dilemma is particularly relevant where legitimation is achieved through argument. How can states argue freely, remaining confident of one another's nonviolent intentions? The need to contain the instability of communicative action is a reminder not to simply assume public spheres are possible in anarchy.

With this in mind I develop the conditions of possibility for communicative action in anarchy. Public interstate talk contains the instability of communicative action. My argument has two elements: a thick notion of international society, and publicity. First, communicative action requires reliable expectations of nonviolence among participants who recognize each another as equals. Providing a snapshot of developments in core institutions of international society—international law, the balance of power, and diplomacy—I argue that, between the Peace of Westphalia in 1648 and the Congress of Vienna in 1814, a horizontal normative order evolved in international politics that organized and regulated the use of violence. But this normative order is not enough. In the second section I therefore develop the role of publicity in the form of face-to-face, multilateral conference diplomacy. Talking in a public forum produces order while keeping the foundations of that order open to rational debate. While

<sup>5</sup> Given the empirical preconditions for nonstate actors to have voice in these sites, only a privileged fraction of world citizens have even this reactive power (see Fine and Smith 2003).

today interstate forums are taken for granted in global governance, in fact this tool was introduced into the system only with the Concert of Europe.<sup>6</sup> I use that case to illustrate the general argument that, in combination with international society, the forum effects of talk sustain international public spheres. Locating the origins of international public spheres in the communicative practices of nineteenth-century autocrats might seem counterintuitive; I defend my use of the case in what follows. The explanatory argument lays the groundwork for a normative claim about the role of horizontal legitimation in global governance, removing the strongest theoretical reason to exclude the states system from global public spheres.

### International Society

The conventional wisdom about the contemporary international system is that it was created by the Peace of Westphalia in 1648, which divided Europe into independent sovereign units. From here, interpretations of anarchy vary considerably. Like many realists in IR, Habermas (1999a, 1999b) treats anarchy as essentially norm-free, and the balance of power as the system's underlying, even natural, logic, making war and strategic competition endemic. But this reflects a historically stunted view of Westphalia. First, the explicit goal of that settlement was to *mitigate* violence. Contemplating the devastation of the Thirty Years War, sovereigns sought better tools to counter drives for continental hegemony. The solution that they hit on, mutual recognition of sovereignty or "anarchy," was an effort to remove religion as a cause of war. Second, after Westphalia, institutions to further regulate violence deepened and became increasingly rationalized. I cannot explore the emergence of international society in detail here (see Osiander 1994), but overall these trends mirrored those Habermas ([1962] 1994b) describes at the domestic level—the differentiation of political practices from an overarching Christian worldview, and a corresponding decline in the role of the sacred. To be sure, as Christian Reus-Smit (1999, 94) points out, for a long time after Westphalia, international institutions retained "premodern" elements with order seen as God given and monarchically protected. Still, institutions adapted in ways sometimes at odds with these values. By 1814, three institutions in particular—international law, the balance of power, and diplomacy—reflected the deepening of a legally constituted horizontal normative order and sphere of nonviolent communication among states.

**International Law.** After Westphalia, what became known later as "international law" became increasingly secularized and anchored in the corporate body of the state rather than individual monarchs or the Church. The religious wars had called into question the

idea of universal Christendom, and after Westphalia "Europe" increasingly replaced it in diplomatic discourse. Secularization was evident in international treaties: religious oaths and references to natural law declined in eighteenth-century legal texts, and states increasingly relied on pragmatic guarantees of various sorts (unilateral, mutual, third party) rather than religious ones (Bull 1977, 33; Satow 1925). There still was a sense of belonging to a "whole" in whose name all diplomacy was aimed, but that whole was increasingly a secular, European "system" whose stability was secured through mutual toleration.

The institution of the state's corporate personality was the basis of the doctrine of *pacta sunt servanda* (sanctity of agreements) which, by obligating the state irrespective of changes in regime, permitted long-term contracting (Anderson 1993, 40; Dunn 1929, 9). Major legal theorists such as Grotius (1583–1645) in the seventeenth-century and Vattel (1714–1767) in the eighteenth-century treated states more than individuals as the core rights-bearing units in the system. In addition, states were increasingly seen as sovereign or autonomous rather than penetrated by other authorities. Juridical autonomy gained ground as the premise of diplomacy and politics. For example, as early as the Utrecht peace negotiations in 1713, precedence concerns were subordinated to pragmatic ones (Osiander 1994, 108). Respect for monarchical supreme authority inside the state was rationalized increasingly through the developing framework of positive law, where law is understood as the will or command of the sovereign backed by threat of sanction. Because by definition no sovereign could be made to obey another, autonomy meant that international law would have to be based on consent. Autonomy also meant that sovereigns retained the exclusive right to judge their own case and thus to take the law into their own hands by waging war (Bull 1977, 28–32; Duchhardt 2000, 283–9). Concentrating the right to act—to contract, sign treaties, wage war—reduced uncertainty about both violence and cooperation, in sharp contrast to medieval structures of overlapping authority and multiple actors.

Importantly, despite being rationalized through divine right, sovereign autonomy was granted to republican states as well as monarchies. This is evident in Utrecht diplomacy, and in Vattel's words, echoed in several legal texts of the eighteenth century: "a dwarf is as much a man as a giant is: a small republic is no less sovereign than the most powerful Kingdom" (cited in Simpson 2004, 32). From there, as Andreas Osiander (1994, 87–8) notes, sovereign "equality was the unavoidable corollary of autonomy. The more there was of the one, the more there had to be of the other." Equality was formally recognized as the basis of diplomacy and international law at the Congress of Vienna.

Furthermore, state practice was becoming the authoritative basis of law, competing with and ultimately replacing the authority of a divine or natural order. A sense coalesced in eighteenth-century legal writings that the states system was a distinct type of social system that operated by its own rules. The legal rules of this system were discovered inductively, through

<sup>6</sup> Among historians, Paul Schroeder (1994) is particularly known for arguing that the Concert of Europe constituted a transformation of European politics. My argument is indebted to his work, although he does not conceptualize the transformation in terms of public spheres

patterns of interaction and treaties. This contrasted both with how law was treated at Westphalia, i.e., mainly as (archaic) custom (Osiander 1994, 48) and with natural law, where rules are deduced from nature and reflect an inherent, universal morality. Although some natural law reasoning can be found in his writings, Vattel in particular is seen to mark the shift toward a law of nations based on consent and state practice (Bull 1977, 33 ff.; Doyle 1992, 269). By the late eighteenth century it was routine to speak in terms of the public law of Europe and "international law" (Suganami 1978).

Some might question whether these developments constitute a normative, much less legal, order. After all, this order eradicated neither war nor dynasticism. The eighteenth century was quite war-prone, and status and succession concerns remained a major cause of war. Moreover, because the evolving legal order had neither legislation nor enforcement, it did not look like law as we understand it in a domestic context. Still, normative order is evident, first, in the fact that wars were subject to rules—engaged in only by states, fought for limited aims, and not fought for religious causes—that limited violence, which distinguished them sharply from the organization of violence in the premodern period. Second, as we saw earlier, a major function served by law is to articulate rules of conduct in terms of the general interest and to convey those rules to subjects. Once law is known, participants are relieved of the burden of constantly negotiating the fundamentals of interaction—who has authority to act, what outcomes can be negotiated; they can fall back on legal norms. From this perspective, international law as it was developing in the eighteenth century certainly served legal functions. Moreover, decision makers and scholars of the period treated international law as law, so that the question of whether it was "really" law never came up. Notions of law as sovereign will coexisted easily with notions of law as rooted in a natural order. The distinction between law, morality, and state political action did not harden until the positivist paradigm consolidated in the nineteenth century (see Vagts and Vagts 1979, 568).

In sum, the trajectory of international law shows that a horizontal normative order took shape in the European states system. The fact that these actors made their power rationalizable according to practice, rather than rank or archaic custom, was the first step toward making it possible for state action ultimately to become subject to public reason.

**The Balance of Power.** It might seem strange to think of the balance of power as an "institution" of international society. Indeed, references in Habermas' writings (1999a, 1999b) suggest that the balance of power operates for him as it does for realists in IR: not as an institution, but mechanically, integrating states through the medium of power, with no normative content. States simply pursue their interests. The system is governed by an equilibrating mechanism, so that no state need deliberately restrain itself or consciously think in terms of a larger interest. This is the bal-

ance of power conceived as invisible hand. But this interpretation of how power operates in international politics overlooks important conceptual and historical aspects of the eighteenth-century balance. It certainly was competitive and war-prone, but not because power was operating anonymously behind sovereigns' backs. Patterns of competition and violence were rooted in and legitimated by shared understandings about the authoritative sources of power and use of force.

As Habermas (1984, 266 ff.) himself argues in the domestic context, although both power and money are media of integration, unlike money, power needs legitimation. Force alone is a brittle source of integration, and Habermas argues that social cohesion ultimately rests on the subjects' felt duty or obligation to submit. His argument is agnostic about the particular legitimating values, i.e., it does not mean that power will be legitimated communicatively or according to standards of reason and equality. The legitimation requirement is general: whatever the prevailing norms, those in power need their rule to be seen as legitimate.

This certainly was true of Europe's balance of power in the eighteenth century, which guided state behavior in ways that were underpinned and rationalized by dynastic and Christian principles. First, the goal of foreign policy—glory—was a reflection of absolutist norms and legitimated competition among sovereigns. War was heroic and associated with ceremony and pageantry, and monarchs looked for opportunities to engage in it in order to achieve glory for the state. As Martha Finnemore (2003, 106–7) puts it, force was a "positive good." Legal norms further sanctioned the sovereign's right to wage war and to declare his own cause just. While monarchs often attempted to negotiate disputes, the fact that norms legitimated sovereign will was a strong incentive to simply act, and to act quickly, which often meant war (Black 1999, 323–5; Gilbert 1951, 7; Hatton 1980, 15).

Second, balancing practices also reflected dynastic norms. Concerns for hierarchy and relative rank among sovereigns meant that there was no norm of trust or cooperation. Alliance loyalties were bargained according to generally accepted rules of "compensation." Any war involved numerous such transactions. Loyalty was not expected; states often were as suspicious of their allies as their adversaries, and indeed often left alliances midwar if proposed a better deal (Finnemore 2003, 105–6; Schroeder 1994). Territory that in the pre-Westphalian period had been seen as held by God's will was now the monarch's property, which allowed it to become a fungible bargaining chip to restore interstate equilibrium (Anderson 1993, 47–8).

Third, despite their struggles for individual glory, the idea of a European balance had normative value for sovereigns: it was their solution to the danger of continental hegemony. Sovereigns agreed that if all pursued equilibrium the continent would remain stable. Thus, beginning with the first modern invocation of a "just equilibrium of power" as the goal for European politics at Utrecht in 1713, actively pursuing balance took on a normative cast. Utrecht negotiators made efforts to link individual goals to the broader systemic goal of



repose or tranquility, and “the European states system was treated as a kind of imaginary super-actor with the same aspirations as the individual actors that made it up” (Osiander 1994, 111). Osiander contrasts this reflexivity about the European states system to the situation at Westphalia, where the parties were simply concerned with restoring the status quo ante and did not discuss Europe as having a distinct identity or needs (*Ibid.* 102). Although the pursuit of glory and the pursuit of balance would seem to be at odds, glory was to be pursued in limited wars for limited aims. The tensions between these aims were not exposed in the system until the Napoleonic wars.

Taken together, these norms manifest a sense of forming a collective, and, as we saw earlier, a shared normative order is a precondition for communicative action to transform power. Here, what we see is a rationalization of the medieval notion of Europe as Christendom into the notion of Europe as a balance-of-power system. That this system was a normative order is clear when considering how Europeans treated outsiders. The boundaries of Europe were culturally determined, and despite its proximity to Europe the non-Christian Ottoman Empire was generally considered outside Europe’s balance-of-power system. Indeed, Christianity was a major legitimating principle in the Concert of Europe vis-a-vis the Ottoman Empire, until 1856. Similarly, what we would now think of as Third World states were not considered members of the balance-of-power system. All of these states were treated by different rules and subject to colonization. Violence was more likely, and less limited, in relations between Europeans and these others. Within Europe, the balance of power mitigated violence; but between Europe and others outside, all bets were off (see Keene 2002; Neumann and Welsh 1991).

**Diplomacy.** Perhaps the most basic precondition for communicative action is that participants can speak to one another without fearing for their lives. That potential evolved among sovereigns in this period. While violence was rife in pre-Westphalian diplomacy, by 1814 European states had pacified the diplomatic sphere.

Several changes helped rationalize interstate communication. First, the consolidation of the state’s corporate agency unfolded at the diplomatic as much as the legal level. The norm of extraterritoriality or diplomatic immunity took root. It became generally accepted that envoys would not be murdered or imprisoned, and weapons not permitted in negotiations (Hatton 1980, 7–8; Langhorne 1981–82, 65–6). Additionally, by 1700 there was a shared understanding that ambassadors officially represented the king. What was called the exchange of formal powers, where diplomats established that they executed policy on the monarch’s behalf, was always the first activity of international conferences (Doyle 1992, 268).

Second, states increasingly defined foreign affairs as a distinct, secular sphere of politics. Medieval diplomacy had reflected the hierarchy of religion over politics. Clergy were ambassadors; Latin was the language

of diplomacy; and the Pope and Holy Roman Emperor had precedence over princes. But the post-Westphalian period was one of bureaucratization, and by the mid-eighteenth century, with states increasingly boasting foreign ministries, there was an educated class of professional bureaucrats with primary loyalty to the state. Ambassadors increasingly were drawn from this group rather than clergy or landed nobles. Professionalization led to a normative shift toward honesty and fair dealing in diplomacy, rather than duplicity. If not always evident in practice, these norms were certainly clear in seventeenth-century manuals on diplomatic method (Anderson, 1993, 46). At the same time, permanent embassies spread; by the Napoleonic Wars, diplomatic communication was virtually continuous among major European capitals.

Third, beginning with Westphalia, an increasingly common practice in foreign affairs was the convening of multistate congresses after wars to construct peace settlements. At first, issues of precedence and method dominated, making congresses difficult to convene and run. Personalized rivalries and protocol disputes consumed inordinate amounts of time and often were settled by duels, even threats of war. But over time, as it became clear that conferences only could proceed once issues of precedence were overcome, references to personal and hierarchical ties declined (Langhorne 1981–82, Nicolson 1954, 42–6).

Diplomacy was further rationalized through the expansion of publicity: the audience for foreign affairs expanded both in reality and in the minds of decision makers. Bureaucratization brought a rise in treaty-printing and record keeping of international events. The new stress on recordkeeping, along with the large delegations that attended conferences, raised the visibility of international politics to those outside the narrow sphere of the king and court (Hatton 1980, 14). Equally important, the eighteenth century saw several publications on international politics meant for a wide audience, from the Abbé de Saint-Pierre’s *Project to Establish Perpetual Peace*, to Rousseau’s vision for European federation, to Kant’s *On Perpetual Peace*. These were not Europe’s first visions of federation, but were noteworthy for their wide dissemination. In addition, both Grotius’ and Vattel’s manuals of international law were widely read among both decision makers and the emerging reading public (Duchhardt 2000, 288–9; Gilbert 1951, 14–5).

Finally, over the eighteenth century an increasing sense developed among statesmen of a “public” below the state whose opinion mattered for diplomacy. Osiander (1994, 104–5) notes that the terminology of public and public opinion was absent from the discourse of Westphalia but figured prominently in negotiations at Utrecht in 1713; and Jeremy Black (1999, 493–4) sees a further rise in this language throughout eighteenth century diplomacy. Whatever the impact of these ideas on foreign policy, their prominence in interstate discourse points to the sense in which diplomacy no longer took place in secret. Foreign affairs were still the realm of princes, and censorship was common, literacy low, and daily newspapers relatively rare in

Europe (Asquith 1978, 102). Still, power politics was generating a critical literature in European civil societies, and sovereigns took note.

In sum, the changes Habermas ([1962] 1994b) identifies in particular states of eighteenth-century Europe as signifying the rise of critical public spheres also had an international dimension. By the Congress of Vienna, developments in international law, balance-of-power thinking, and diplomacy had given rise to a horizontal normative order constituted by mutual recognition of sovereignty and ongoing communication.

## Interstate Public Spheres

A normative order organizing the use of violence is an important accomplishment but alone cannot sustain public sphere governance. Because the institutions of international society do not reliably hold violence at bay, they cannot backstop political argument the way centralized authority does. But rather than look to exogenous sources of nonviolence like the democratic peace, I propose a source internal to the states system: publicity. Public spheres require that power holders' actions are visible and that those affected can deliberate and form opinions about that power. The necessary visibility has two dimensions: face to face (horizontal) and the more mediated visibility produced by communications technologies (vertical). Extant theory stresses the latter. My focus is the former: when states met in 1814 to discuss the European balance of power, this new practice—conference diplomacy—introduced the power of face-to-face publicity into the states system in a systematic way. I propose that face-to-face publicity generates forum effects of talk, which help keep violence at bay and make possible public sphere governance. The forum effects permit international society's norms to become more salient in interstate decision making, even where states contemplate the use of force. This is as evident in the Concert of Europe, with which I illustrate the argument, as in contemporary multilateralism.

**The Forum Effects of Talk.** <sup>1</sup> My argument begins by specifying an action context, the forum, as the arena of interstate talk. Forums have two salient characteristics. First, discussions within them are premised on nominal equality. This ensures that all have the same right to speak and to be heard. Second, forums are public: they consist of more than two participants meeting face to face, and outsiders are aware of the meetings. Roles such as publicist and reporter, and media such as minutes of meetings, pamphlets, newspapers, television, radio, and so on, make discussions visible to a broad audience outside the decision-making context.

The proposed effects of forum publicity depend on the motivational assumption that actors care how they appear to others. This thin assumption does not rely on altruism among speakers; but it does not refer to caring for one's reputation in a rationalist sense. That is, actors do not necessarily care how they appear because there are future benefits to gain or material costs to suffer from appearing in a certain way. Rather the assumption

is that part of what it means to be a social actor is to care what others think of you, and this is made manifest in the forum.

From here, the argument is that publicity has three "forum effects." First, drawing on Jon Elster's work (1995), I argue that when in public even selfish actors will want to appear impartial and fair and so will generalize their interest claims and argue impartially (also see, e.g., Lynch 1999; Risse 2000; Schimmelfennig 2001). For example, "This is in England's interest," would become "This is a great power interest," or "a matter of sovereign equality." Selfishness expressed in public must be rendered in terms acceptable to all. Moreover, when states speak impartially in public, they can find themselves subsequently compelled to follow through on commitments based on those rationales. What Elster calls the "civilizing force of hypocrisy" can thus lead to more equitable group outcomes than if powerful actors did not have to justify their actions in public. Indeed, studies in a variety of disciplines support the claim that face-to-face talk has beneficial effects on joint problem solving (e.g., Ostrom 2000).

The other forum effects develop over time. With continued expectations that they will meet in forums, speakers get habituated to practices of reason giving and to relying on public criteria of acceptability. These habits are effects of the public context: speakers see themselves as acting less as "selves" than as "members" of a group. Over time, habits acquire normative weight, translating to the second and third forum effects: a norm of publicity develops, by which I mean a procedural reciprocity where participants feel they must make their reasons available to others; and public reason develops, by which I mean that the general and impartial arguments they regularly invoke increasingly are seen as shared norms. Public reason becomes a collective belief structure, a shared sense of the "right" reasons for action: a public sphere.

My claim that the forum effects of talk help generate public spheres explicitly links order to legitimation. On the one hand, the forum effects help produce order by dampening the security dilemma. As foreign policy behaviors become defined and justified similarly by all participants, states have greater certainty regarding what problems are, and about what actions will be ignored and which might be sanctioned. By setting the parameters of conflict, the forum effects of talk steer interstate competition in a way to buffer the group against the most disastrous outcomes. That is, they cause self-restraint.

At the same time, interstate argument opens up the possibility for public reason to legitimate international outcomes. In one sense this is a habituation argument: a sociological norm to give reasons develops among actors who recognize one another as nominally equal; these actors need not be democrats, and they need not care about legitimation. But the fact that the habit is one of reason giving links the forum effects to the democratic intent of public sphere theory, making possible what Risse (2000) calls a "logic of arguing." Where justifications for action are public and coalesce around notions of a general, impartial interest, it becomes

possible for participants and the audience to link the state's rhetoric to their own internal morality, and even to take the communicative orientation necessary for public reason and communicative consensus in a Habermasian sense. More generally, in the context of ongoing interstate public spheres, public claims can ultimately evolve to commitments to justice if outsiders, participants, and later generations affirm them as such. Interstate public spheres thus establish conditions of possibility for communicative action in world politics.

In short, in many ways interstate publics are like decision-making publics (or parliaments) inside states. Both involve managing joint problems through collective decisions. In both, discussion is aimed at consensus and the consensus achieved is expected to be binding. In both, participants are expected to remain engaged in discussion; each "consensus" is only a contingent resolution to a problem, not the end of the discussion. Finally, both serve as focal points of critical public spheres.

The difference is that domestically, formal institutions backed by the state's coercive power create the decision-making public and guarantee that collective decisions will be implemented. Among states, in contrast, the expectation of binding consensus is not enforceable and sometimes not even institutionalized, but must be sustained by the public sphere discussion itself. States engaged in joint problem solving never actually cede their own (nominal) authority to act. As such, interstate public spheres do not guarantee an end to war. States can always destroy the conversation or render it meaningless by exiting and resorting to violence. However, as the expectation to keep talking grows, the sense that any consensus is binding grows both among participants and in the broader audience of their deliberations, even without coercive guarantees. And the exit option can become less attractive insofar as states recognize their interdependence and realize that unless each stays at the table, all will suffer. As such, the forum effects of talk can make it possible over time to domesticate certain problems completely, or at least bring them out of the realm where resort to violence is routine.

**The Concert of Europe.** The first case of conference diplomacy in Europe illustrates how public talk can contain the instability of communicative action. The 1814 Vienna Settlement after the Napoleonic Wars introduced the practice of face-to-face consultation as a strategy for managing the balance of power. In the early post-Vienna years, the European Great Powers—Great Britain, Austria, Prussia, Russia, and France—all faced domestic unrest and revolutions in Spain, Naples, Portugal, and Greece. The French Revolution had demonstrated the severe threat liberal revolution could pose to the balance of power, and the Great Powers all felt a special responsibility to prevent that from happening again. At the same time, they did not agree on how to prevent revolutions from spreading, or on who should benefit, which meant Great Power war remained a possibility. In this context, the Great Powers met in a spurt of congresses from 1819

to 1822. These public forum discussions made a difference: when unilateral action did occur, such as Austria in Naples, Russia against the Ottoman Empire, it was limited and nonexpansionist.

In my view, Concert self-restraint cannot be understood separately from the practice of conference diplomacy. These former rivals were able to cooperate publicly when they could not privately. To be sure, the Concert of Europe still retained important aspects of the eighteenth-century system: a balance of power constituted primarily by absolute monarchs. But the visibility provided by conference diplomacy introduced a new, horizontal dynamic of publicity. Meeting face-to-face made the balance of power visible to those who constituted it, and this made a difference. Although a full case study is beyond the scope of this paper (see Mitzen 2001), a brief example of Concert governance lends support to my "one-step" hypothesis that public interstate talk can produce order and legitimation simultaneously.

A central problem the Concert Powers faced in the 1820s was the decline and potential break-up of the Ottoman Empire—the "sick man of Europe"—which was not a member of the Concert and whose decay could lead to Great Power conflict over the spoils, which was routine in the eighteenth century. The Greeks, who had been under Ottoman rule for hundreds of years, precipitated a crisis by rebelling in 1821. The revolt lasted several years, and the resulting Balkan instability threatened to erupt into Great Power war. Russia was the power to watch: it had grievances against the Ottomans, sympathy for the Greeks as fellow Orthodox Christians, and the most to gain materially by intervention. In 1821, no Great Power knew what to do. None wanted Great Power war, not even Russia, but it was unclear how to avoid it. In this situation, realists would predict Great Power war; and statesmen themselves all expected it. Yet it did not happen.

The Great Powers avoided war over the Greek revolt by publicly "Europeanizing" the Greek Question, that is, by adopting a collective definition of the Greek revolt that kept it within the parameters of their pre-existing cooperation. This was not easy: the Ottoman Empire had not signed the Vienna Settlement and was not considered a sovereign the way European states were considered sovereign. Additionally, it was not obvious at the time that the Greeks were Europeans who deserved to be under the Concert's purview. In short, violence in the Balkans was essentially an "out of area" problem. In the midst of this uncertainty, public talk made a difference. Invoking Greece as a European problem in diplomatic conferences created a discursive structure, which regulated Great Power choices in a way that private diplomacy could not, and made it possible to solve the Greek Question without Great Power war. One might say that the Great Powers "talked Greece into Europe."

**1821–3.** Managing the Greek Revolt had two phases, 1821–3 and 1826–32, and in both the crucial concern was to prevent Russian intervention on behalf of the

Greeks, which all felt would escalate into Great Power war. In 1821–3, the Concert strategy was to interpret the Balkan revolt as part of the epidemic of liberal revolutions sweeping Europe. On that basis they proposed that the Concert side with the legitimate sovereign (in this case, the Ottoman Sultan) against the Greeks. The Great Powers made these arguments both publicly and privately; but the arguments restrained Russia only when made publicly. More specifically, the Greek revolt broke out while the Great Powers were in the midst of the Laibach Congress, which had been convened to address a different revolt in Naples and at which they had recently decided to support the Neapolitan sovereign against the revolutionaries. Faced with this new revolt, the Great Powers forged an initial consensus that it was part of the same “European conspiracy” against European thrones and needed to be quashed. A joint allied declaration against the revolt was publicized immediately. Without support, it petered out (Schroeder 1994, 610 ff.).

But soon after Laibach, another Greek revolt broke out. Unlike the first, which had been relatively small scale, this one engaged every stratum of the Greek population, from clergy to nobility to peasants, and quickly gathered momentum. The Ottomans responded with hard-line measures, such as hanging Greek clergy and massacring Christians. With the Laibach Congress no longer in session, Russia began to assert its pro-Greek interests and talk of unilateral intervention. Each Great Power tried private diplomacy to restrain Russia, using the same cognitive frame—this was a liberal revolt against a legitimate sovereign—they had used at Laibach (Kissinger 1957, 293–4). But private diplomacy did not work. Prussia and France took actions that seemed to reflect unsteady support of the Laibach interpretation of Greece, while British and Austrian intentions were suspect (Anderson 1966, 58). Private diplomacy made it difficult to “see” the collective European interest in supporting Turkish sovereignty over the Greeks, generating uncertainty: uncertainty about the rules of the game that applied to the Balkans, uncertainty about their own and each other’s interests, and uncertainty about what to do. Fears of Great Power war intensified.

Restraining Russia became possible, however, when Britain and Austria adopted a public strategy. Their diplomacy had the same cognitive components—it supported the Ottoman sovereign against Greek rebels—but it was newly public. Although the strategy was spearheaded by Britain and Austria and not the Concert as a whole, the two states took care to ensure that their bilateral meetings were not seen by Russia as a budding counteralliance, for example, by choosing not to issue a joint communiqué condemning Russia. They also called for a congress specifically on the Greek Question, and informed Prussia and France of the congress and its rationale, to appear as a united front. The combination of drawing on publicly accepted arguments and linking those arguments to a public forum involving the entire alliance meant that, from then on, the Greek question was addressed as a general interest. That this strategy indeed restrained Russia is

evident in 1822, where Metternich persuaded Russia not to intervene by invoking the upcoming Congress, and then the Concert ratified its stance publicly at the Congress. The war party in Russia was silenced; all Greek members of the Russian diplomatic corps subsequently resigned or were purged (Anderson 1966, 61 ff.; Nichols 1961, 55 ff.).

My claim is that this collective, public strategy worked because it made the European interest in stability visible to one another and to Russia, which reduced uncertainty and provided a concrete referent for that collective interest, the forum. Whatever any leader thought privately about Greece or the Ottoman Empire, appearing in public kept the collective interest salient for all of them, which caused self-restraint.<sup>7</sup>

**1826–32.** Despite the initial Great Power success, the Greek revolt persisted. As the decade wore on it was increasingly clear this was not a liberal revolution, the original concern of the Concert, and that the decline of Ottoman sovereignty posed a different sort of threat to the European balance of power. The Great Powers still felt that somehow the Greek Revolt was “their” problem, and so later in the decade they turned again to the Concert forum. This time the collective belief that the Greek revolt posed a European question was made concrete through the 1827 Treaty of London, which committed the Great Powers to resolving the Greek Question jointly and without war. The Treaty of London did not prevent war altogether: Russia did fight the Ottomans in 1828. But Russia’s justifications for that war had nothing to do with Greece and its war aims were limited. What the Treaty did was help prevent a war between Russia and the other Great Powers over Greece. In the war, Russian troops inched down through the Balkans. Security dilemma logic tells us that a larger war could easily have been triggered by other Great Powers fearing Russian expansion. Invoking the Treaty of London gave the Concert powers, including Russia, a common reference point, and a public one, for their joint commitment to the European status quo and to keep the Balkan issue separate. By virtue of its public commitment, in other words, Russia restrained itself: the war remained limited (Jelavich 1991, 84 ff.).

Keeping the war contained enabled Great Power governance. The London Conference on Grecian Affairs (1827–32), an ongoing conference at the ambassadorial level and the first of its kind, was set up to solve the Greek Question once and for all. The ambassadors negotiated a French occupation of the Greek mainland, and the constitution, frontiers, population, and even king of the new state. Such a thing—jointly midwifing the birth of a nation-state—had never been done before.<sup>8</sup> On top of that, here it was done deliberately: proposals were put forward and debated out of the

<sup>7</sup> This claim would be contested by, e.g., Rendall 2000.

<sup>8</sup> “The” Conference was actually several meetings at the ambassadorial level of Treaty of London signatories. In the primary documents each meeting is referred to as a separate conference, but since the same actors engaged in discussions and the meetings fell under a single mandate from the Treaty of London it became common to

heat and light of high politics. Because the negotiators did not constantly have to keep their eye on Russia they could freely discuss the problem. Moreover, the minutes and final protocols were made public, and were referred to by the Great Powers in the war diplomacy. Invoking the protocols helped keep the Greek Question out of the war.

In sum, it seems clear that the forum made a difference. Surely if the Great Powers had wanted war there would have been war. But even when states do not want war the security dilemma tells us war still can happen. Repeatedly in the 1820s, the Concert of Europe forum provided a concrete reference point—a publicly shared commitment to the collective interest in peace. Without the Treaty of London, war would have been more likely; without the London Conference there would have been no Greek independence. Neither outcome can be understood without incorporating the dynamics of public talk.

**An International Public Sphere?** It is possible to grant my argument that Concert publicity helped prevent war but to reject the notion that those dynamics constituted anything like a “public sphere.” Certainly the values driving Concert cooperation are at odds with those of public sphere theory. Only one Great Power, Britain, was a democracy, and its democracy was quite limited, whereas a crucial goal of Britain’s Concert partners was to defend monarchy and prevent liberal revolution. Still, three aspects of this case suggest the applicability and importance of public sphere theory.

First, Concert diplomacy introduced a new, intersovereign visibility to European interstate politics. Sovereigns who were accustomed to making foreign policy unilaterally and in secret suddenly found themselves justifying their policies to fellow sovereigns. This was different from how diplomacy had been practiced in the eighteenth century and the Napoleonic Wars, and signified, as Paul Schroeder (1994) seminally argues, a “transformation” of European politics. Of course, the Concert was by no means a realization of the public sphere ideal. Its diplomacy did not embody (or attempt to embody) Habermasian ideals of free publicity and rational communication. But in fairness, no actually existing public sphere today fully embodies these ideals. Power and privilege always matter and are problems even domestically. The importance of the public sphere concept is as a guide, to determine whether and how a given exercise of power is normatively better or worse than another. In my view, public sphere theory helps us make sense of the dynamics that conference diplomacy set in motion as normatively improved action. The Concert’s goal was functional: avoid Great Power war. But where power is called on to give reasons to a relevant public, if the preconditions exist, and I have argued that they did at the intersovereign level, then reason giving can have effects. Injecting horizontal

publicity into their foreign policy decision making, in an environment where vertical publicity similarly was on the rise, set in motion processes capable of transforming political authority.

Second, Great Power diplomacy took on characteristics analogous to those of a decision-making public in the domestic context. Participants saw themselves as charged with common problems and deliberated about them in conferences premised on nominal equality, publicity, and the goal of consensus. Indeed, there is evidence that, despite their autocratic domestic governments, Metternich and Alexander, who were schooled in Kant, were committed to deliberating with each other and to the idea of European federation of sorts, even if in a qualified sense and for instrumental reasons (Kann 1960; Sofka 1998). The impulse to dismiss intersovereign communication as “nonpublic” because it is engaged in by sovereigns rather than citizens of liberal societies overlooks the real power publicity had in this case. After all, domestic public spheres are not solely characterized by vertical dynamics; the horizontal decision-making dimension is crucial. Overall, the horizontal dimension has received considerably less attention in public sphere theory and may even be one reason for skepticism of my claims about this new diplomatic practice. Recognizing horizontal publicity as a dimension of all public spheres, it becomes easier to accept in principle that the public speech among decision makers who are constituted as equals and are aware that they are being watched potentially has civilizing effects. While Concert diplomacy was dominated by elites who were not fully committed to the vertical dimension of publicity, their diplomacy was “public” in the limited sense that these leaders were committed to justifying their policies to one another. They were vetting their decisions through each another—equals who were affected by the use of state power.

Third, the fact that Concert decision makers were aware of a literate European public who knew about the conferences injected the beginnings of the vertical dynamics we associate with contemporary critical public spheres. Part of my argument is that in the Concert period a critical public was aware of and commenting on foreign affairs, which influenced the changes in diplomatic practice in a process that might be seen as an international corollary to the processes Habermas (1994b) chronicles. Although it was difficult for European public opinion to affect decision makers in this period, there is evidence that from 1821 to 1827 public opinion helped keep the Greeks’ European, Christian identity salient in decision makers’ minds. The revolt received enormous press coverage. Outside Concert forums, pro-Greek societies formed and thrived and volunteers flocked to the cause. The widespread feeling that Greeks were a piece of Europe under the rule of the barbarian, infidel “Turk,” made it difficult for the Great Powers to ignore the revolt. In addition, the steady message that the Greeks were Europeans enabled enterprising Greek leaders to earn the right to have a say in their political outcome. No prior nonstate rebel group had earned this right—not the

refer to them in retrospect as a single conference (see Dunn 1929). The London conference was closely followed by an ambassadorial conference on the Belgian question, which also created a sovereign state (Schroeder 1994, 670–91).

Serbs, not the Romanians, and certainly not the Poles. By contrast, even as the Greeks' war fortunes fell their international legal personality grew, enabling them ultimately to appeal to the Great Powers for recognition (Cunningham 1978).

It is important to view the Concert through a public sphere lens because this was the first case of multilateral governance and in it we see seeds of the two dimensions of public spheres that are in full flower today. To ignore the public sphere dimension of the Concert story is to overlook what was so innovative about Concert diplomacy, namely the practice of public consultation, which, in the context of international society and rising critical publicity, enabled public reason for the first time to guide state decisions about the use of force.

Of course, today global governance is everywhere, and I do not mean to suggest that the forum effects of talk are the only or even the main cause of the increased ability to cooperate. The spread of democracy and global capitalism clearly are important as well. But the forum effects should not be overlooked. In the domestic case, the fact that centralized enforcement guarantees public spheres does not mean it causes every governance outcome. Indeed, speakers tend to be so habituated to deliberation that they hardly notice that their right to speak is coercively protected. But at times deep normative disagreements arise, and it is possible to continue discussion only by invoking the threat of the state. Arguably, the forum effects of talk function similarly at the international level: in the breach, where normative disagreement threatens global governance, they preserve the potential for joint problem solving.

## CONCLUSION

The instability of communicative action makes anarchy a hard case for public spheres. Where legitimation is accomplished through argument, future disagreement must always be possible, but it must not devolve into violence that can destroy the social order. The global governance two-step obscures this connection between order and legitimation, by locating global public spheres in already orderly environments and relegating states system dynamics to order production alone. In contrast, I have proposed a one-step theory that links order and legitimation. Multilateral diplomacy constitutes the horizontal dimension of global public spheres. More specifically, forum discussion among states mitigates the problem of violence by generating a structure of public reason. Public reason channels outcomes while keeping the rationales for action open to debate.

One way to draw out the implications of my argument is to consider how it would make sense of a contemporary case such as the diplomatic run-up to the Iraq War. Perhaps most importantly, it makes sense of what participants themselves thought they were doing when they engaged in rounds of public debate. They were determining whether the use of force was the right thing to do. The United States offered several justifications: international law against weapons

of mass destruction, democratic norms, failure of the UN sanctions regime, and self-defense; yet it could not forge consensus. Although Security Council members agreed that Iraq had violated international law and that the sanctions regime was not working, they disagreed that it posed an imminent threat to international peace and security and rejected the justification of self-defense.

Stepping back from the practitioner's to an analyst's perspective, my framework yields a distinct point of view on debates over the fact that the United States acted anyway. Explanatory analysts might see the United States' turn away from the UN and reliance on a "coalition of the willing" as a failure of the global public sphere: in the breach, it could not restrain the system's most powerful state. Diplomatic talk proved cheap. Indeed, the case does seem to suggest a breakdown of the public sphere in that no consensus was reached and the United States acted anyway.

However, the UN debate served three important functions. First, consider the counterfactual: what if the United States had not brought its case to the UN and engaged in public talk, instead simply unilaterally deposing Saddam Hussein? It is hard to imagine other Great Powers would not have used this as permission for their own unilateralism, perhaps, for example, in Chechnya or Taiwan. Even if talk cannot always prevent the powerful from acting, a salient consensus can deter some potential law-breakers. Silence, in contrast, intensifies the security dilemma, making it more difficult to convert hard, military balancing into soft, diplomatic balancing. Second, public talk changed how the United States pursued its interests, delaying the use of force and making it necessary to act through the "coalition of the willing" rather than unilaterally. Finally, the lack of consensus is proving costly. It may be too soon to assess the full impact of international criticism, but the United States clearly already has lost social capital in the international community.

Other analysts, from a normative perspective, might highlight the tension between legality and morality raised by the fact that, regardless of whether the war is seen as legal, the end result may just be a democratic Iraq. Reactions to the 1999 Kosovo intervention are instructive in this respect. There, theorists argued although the United States and NATO acted illegally by not gaining UN approval, the use of force was justifiable after the fact, in the name of the cosmopolitan morality of human rights (e.g., Buchanan 2001; Habermas 1999a). Global morality, premised on the intrinsic worth of individuals, is outpacing international legality, premised on state sovereignty; and in this transition it is perhaps necessary to tolerate violations of international law in the name of the emerging moral order. On this view, consensus at the UN itself is not intrinsically important, if human rights violations are sufficiently severe.

Extrapolating to the Iraq case, this suggests some difficulties of condemning the war from a cosmopolitan perspective. Saddam Hussein was a major



human rights violator who abused his population for decades. If Iraqi citizens gain significant rights protections as a result of the intervention, then the fact that the United States bypassed the UN could perhaps be overlooked. As long as right-thinking states do the right thing—promote human rights, even if it sometimes means violating sovereignty or bypassing international procedures—then force can be justified.

My framework cautions against such a “turn to ethics” (Koskeniemi 2002) in international law. To favor morality over legality gives liberal norms, as interpreted by advanced industrial democracies, pride of place. This is in effect a withdrawal of the Westphalian permission to remain strangers. The dangers are both that so-called pariah states whose values are not legitimated might withdraw altogether from the conversation, and that would-be imperialists could feel justified in doing the same. In a world that is silent across borders, the balance of power becomes once again an invisible hand. This brings us back to the eighteenth century, or forward to a clash of civilizations (Huntington 1993). Devalorizing public attempts to achieve interstate consensus makes anarchy a more dangerous place.

In sum, global democracy is an even greater challenge than we thought. It must balance needs for democracy at both horizontal and vertical levels rather than allowing either sphere to triumph. The fact that reasons that win in international forums must prioritize democratic values does not preclude nondemocracies from participating in global public spheres. In the end, we may want to make cosmopolitan arguments about how governance should evolve in the states system; we may even theorize international public spheres in the explicit hope of helping cosmopolitan values triumph. Visualizing interstate public spheres is not meant to diminish the importance of domestic-level democracy or transnational civil society. The point is to highlight how, for nearly 200 years, international institutions and publicity have helped regulate violence and how they have injected into anarchy the possibility of public reason in world politics.

## REFERENCES

- Alvarez, José. 2001. “Do Liberal States Behave Better? A Critique of Slaughter’s Liberal Theory.” *European Journal of International Law* 12 (April): 183–246.
- Anderson, M. S. 1966. *The Eastern Question, 1774–1923. A Study in International Relations*. New York: St. Martin’s.
- Anderson, M. S. 1993. *The Rise of Modern Diplomacy 1450–1919*. New York: Longman Group.
- Asquith, Ivon. 1978. “The Structure, Ownership, and Control of the Press, 1790–1855.” In *Newspaper History. From the 17th Century to the Present Day*, ed. George Boyce, James Curran, and Pauline Wingate. Thousand Oaks, CA: Sage Publications, 98–116.
- Black, Jeremy. 1999. *Eighteenth-Century Europe*. 2nd ed. New York: St. Martin’s.
- Bohman, James. 1994. “Complexity, Pluralism, and the Constitutional State: On Habermas’ *Faktizität und Geltung*,” *Law & Society Review* 28 (4): 897–930.
- Bohman, James. 1997. “The Public Spheres of the World Citizen” In *Perpetual Peace: Essays on Kant’s Cosmopolitan Ideal*, ed. James Bohman and Matthias Lutz-Bachmann. Cambridge, MA: MIT Press, 179–200.
- Bohman, James. 1999. “International Regimes and Democratic Governance: Political Equality and Influence in Global Institutions” *International Affairs* 75 (July): 499–513.
- Brunkhorst, Hauke. 2002. “Globalising Democracy without a State: Weak Public, Strong Public, Global Constitutionalism.” *Millennium* 31 (3): 675–90.
- Brunnée, Jutta, and Stephen J. Toope. 2000. “International Law and Constructivism: Elements of an Interactional Theory of International Law.” *Columbia Journal of Transnational Law* 39 (19): 19–73.
- Buchanan, Allen. 2001. “From Nuremberg to Kosovo. The Morality of Illegal International Legal Reform.” *Ethics* 111 (July): 673–705.
- Bull, Hedley. 1977. *The Anarchical Society*. New York: Columbia University Press.
- Cederman, Lars-Erik. 2001. “Back to Kant: Reinterpreting the Democratic Peace as a Macro-Historical Learning Process,” *American Political Science Review* 95 (March): 15–32.
- Checkel, Jeffrey. 2001. “Why Comply? Social Learning and European Identity Change.” *International Organization* 55 (Summer): 553–88.
- Cochran, Molly. 2002. “A Democratic Critique of Cosmopolitan Democracy.” *European Journal of International Relations* 8 (December): 517–48.
- Cunningham, Allan. 1978. “The Philhellenes, Canning and Greek Independence.” *Middle Eastern Studies* xiv: 151–81.
- Doyle, Michael. 1986. “Liberalism and World Politics.” *American Political Science Review* 80 (December): 1151–69.
- Doyle, William. 1992. *The Old European Order 1660–1800*. 2nd ed. New York: Oxford University Press.
- Duchhardt, Heinz. 2000. “War and International Law in Europe Sixteenth to Eighteenth Centuries.” In *War and Competition between States*, ed. Philippe Contamine. Oxford: Clarendon Press, 279–99.
- Dunn, Frederick Sherwood. 1929. *The Practice and Procedure of International Conferences*. Baltimore, MD: Johns Hopkins Press.
- Ellis, Jaye. 2002. “International Regimes and the Legitimacy of Rules: A Discourse-Ethical Approach.” *Alternatives* 27 (July–September): 273–300.
- Elster, Jon. 1995. “Strategic Uses of Argument.” In *Barriers to Conflict Resolution*, ed. Kenneth Arrow et. al. New York: W. W. Norton, 236–58.
- Erikson, Erik Oddvar, and John Erik Fossum, eds. 2000. *Democracy in the European Union: Integration through Deliberation*. New York: Routledge.
- Fine, Robert, and Will Smith. 2003. “Jurgen Habermas’s Theory of Cosmopolitanism.” *Constellations* 10 (4): 469–87.
- Finnemore, Martha. 2003. *The Purpose of Intervention*. New York: Cornell University Press.
- Flyvbjerg, Bent. 1998. *Rationality and Power*. Chicago: University of Chicago Press.
- Fraser, Nancy. 1992. “Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy.” In *Habermas and the Public Sphere*, ed. Craig Calhoun. Cambridge, MA: MIT Press, 109–42.
- Gilbert, Felix. 1951. “The ‘New Diplomacy’ of the 18th Century,” *World Politics* 4 (October): 1–38.
- Habermas, Jürgen. 1984. *Theory of Communicative Action*, vol. 2. Boston: Beacon Press.
- Habermas, Jürgen. 1994a. *The Past as Future*. Lincoln, NE: University of Nebraska Press.
- Habermas, Jürgen. [1962] 1994b. *Structural Transformation of the Public Sphere*. Cambridge, MA: MIT Press.
- Habermas, Jürgen. 1996. *Between Facts and Norms*. Cambridge, MA: MIT Press.
- Habermas, Jürgen. 1997. “Kant’s Idea of Perpetual Peace, with the Benefit of 200 Years’ Hindsight,” In *Perpetual Peace: Essays on Kant’s Cosmopolitan Ideal*, ed. James Bohman and Matthias Lutz-Bachmann. Cambridge, MA: MIT Press, 113–53.
- Habermas, Jürgen. 1998. “The European Nation-State: On the Past and Future of Sovereignty and Citizenship.” In *The Inclusion of the Other: Studies in Political Theory*, ed. Ciaran Cronin and Pablo de Grieff. Cambridge, MA: MIT Press, 105–27.
- Habermas, Jürgen. 1999a. “Bestiality and Humanity. A War on the Border between Legality and Morality.” *Constellations* 6 (3): 263–72.

- Habermas, Jürgen. 1999b. "A Short Reply." *Ratio Juris* 12 (December): 445–53.
- Habermas, Jürgen. 2001. *The Post-National Constellation*, ed. Max Pensky. Cambridge, MA: MIT Press.
- Hatton, Ragnhild. 1980. "Nijmegen and the European Powers." In *The Peace of Nijmegen 1676–1679*, ed. J. A. H. Bots. Amsterdam: Holland University Press, 1–17.
- Huntington, Samuel. 1993. "The Clash of Civilizations." *Foreign Affairs* 72 (Summer): 22–49.
- Jelavich, Barbara. 1991. *Russia's Balkan Entanglements 1806–1914*. New York: Cambridge University Press.
- Johnston, Alastair Iain. 2001. "Treating International Institutions as Social Environments." *International Studies Quarterly* 45 (December): 487–515.
- Johnstone, Ian. 2004. "US-UN Relations after Iraq: The End of the World (Order) as We Know It?" *European Journal of International Law* 15 (December): 813–38.
- Kann, Robert A. 1960. "Metternich: A Reappraisal of his Impact on International Relations." *The Journal of Modern History* 32 (December): 333–39.
- Keene, Edward. 2002. *Beyond the Anarchical Society*. New York: Cambridge University Press.
- Keohane, Robert. 1984. *After Hegemony*. Princeton, NJ: Princeton University Press.
- Kissinger, Henry A. 1957. *A World Restored: Metternich, Castlereagh and the Problems of Peace 1812–22*. Boston: Houghton Mifflin Co.
- Koskenniemi, Martti. 2002. "'The Lady doth Protest too Much': Kosovo, and the Turn to Ethics in International Law." *The Modern Law Review* 65 (March): 159–75.
- Krasner, Stephen, ed. 1983. *International Regimes*. Ithaca, New York: Cornell University Press.
- Kratochwil, Friedrich. 1989. *Rules, Norms, Decisions: On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs*. New York: Cambridge University Press.
- Langhorne, Richard. 1981–82. "The Development of International Conferences, 1648–830." *Studies in History and Politics, Special Issue: Diplomatic Thought 1648–1815*, 2: 61–75.
- Legro, Jeffrey. 1996. "Culture and Preferences in the International Cooperation Two-Step." *American Political Science Review* 90 (March): 118–37.
- Linklater, Andrew. 1998. *The Transformation of Political Community*. Cambridge: Polity Press.
- Lynch, Marc. 1999. *State Interests and Public Spheres: The International Politics of Jordan's Identity*. New York: Columbia University Press.
- Mitzen, Jennifer. 2001. "Toward a Visible Hand: The International Public Sphere in Theory and Practice." Ph.D. diss. University of Chicago.
- Mitzen, Jennifer. n.d. "Ontological Security in World Politics." Manuscript, Ohio State University.
- Müller, Harald. 2001. "International Relations as Communicative Action." In *Constructing International Relations*, ed. Karin Fierke and Knud Erik Jørgensen. New York: M. E. Sharpe, 160–78.
- Neumann, Iver B., and Jennifer Welsh. 1991. "The Other in European Self-Definition." *Review of International Studies* 17: 327–48.
- Nichols, Irby C., Jr. 1961. "The Eastern Question and the Vienna Conference, September 1822." *Journal of Central European Affairs* 21 (April): 53–66.
- Nicolson, Harold. 1954. *The Evolution of Diplomatic Method*. New York: Macmillan.
- Osiander, Andreas. 1994. *The States System of Europe 1640–1990*. Oxford: Clarendon Press.
- Ostrom, Elinor. 2000. "Collective Action and the Evolution of Social Norms." *Journal of Economic Perspectives* 14 (Summer): 137–58.
- Payne, Rodger. 2001. "Persuasion, Frames, and Norm Construction." *European Journal of International Relations* 7 (March): 37–61.
- Rendall, Matthew. 2000. "Russia, the Concert of Europe, and Greece, 1821–29. A Test of Hypotheses about the Vienna System." *Security Studies* 9 (Summer): 52–90.
- Reus-Smit, Christian. 1999. *The Moral Purpose of the State*. NJ: Princeton University Press.
- Risse, Thomas. 2000. "'Let's Argue!' Communicative Action in World Politics." *International Organization* 54 (Winter): 1–39.
- Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97 (November): 585–602.
- Rubin, James P. 2003. "Stumbling into War." *Foreign Affairs* (September/October): 46–66.
- Ruggie, John G. 1998. "What Makes the World Hang Together? Neo-Utilitarianism and the Social Constructivist Challenge." *International Organization* 52 (Autumn): 855–86.
- Samhat, Nayef, and Rodger Payne. 2003. "Regimes, Public Spheres, and Global Democracy." *Global Society* 17 (July): 273–95.
- Satow, Ernest. 1925. "Pacta Sunt Servanda or International Guarantee." *Cambridge Historical Journal* 1 (3): 295–318.
- Schimmelfennig, Frank. 2001. "The Community Trap: Liberal Norms, Rhetorical Action, and the Eastern Enlargement of the EU." *International Organization* 55 (Winter): 47–80.
- Schroeder, Paul. 1994. *The Transformation of European Politics, 1763–1848*. New York: Oxford University Press.
- Simpson, Gerry. 2004. *Great Powers and Outlaw States*. New York: Cambridge University Press.
- Slaughter, Ann-Marie. 1995. "International Law in a World of Liberal States." *European Journal of International Law* 6: 503–38.
- Sofka, James R. 1998. "Metternich's Theory of European Order: A Political Agenda for 'Perpetual Peace.'" *The Review of Politics* 60 (Winter): 115–49.
- Suganami, Hidemi. 1978. "A Note on the Origin of the Word 'International.'" *British Journal of International Studies* 4: 226–32.
- Vagts, Alfred, and Detlev F. Vagts. 1979. "The Balance of Power in International Law: A History of an Idea." *American Journal of International Law* 73 (October): 555–80.
- Wendt, Alexander. 1999. *Social Theory of International Politics*. New York: Cambridge University Press.
- Wendt, Alexander. 2004. "The State as Person in International Theory." *Review of International Studies* 30 (April): 289–316.
- White, Stephen K. 1994. *The Recent Work of Jürgen Habermas*. New York: Cambridge University Press.

# Prenegotiation Public Commitment in Domestic and International Bargaining

BAHAR LEVENTOĞLU *Stony Brook University*

AHMER TARAR *Texas A&M University*

**W**e use a formal bargaining model to examine why, in many domestic and international bargaining situations, one or both negotiators make public statements in front of their constituents committing themselves to obtaining certain benefits in the negotiations. We find that making public commitments provides bargaining leverage, when backing down from such commitments carries domestic political costs. However, when the two negotiators face fairly similar costs for violating a public commitment, a prisoner's dilemma is created in which both sides make high public demands which cannot be satisfied, and both negotiators would be better off if they could commit to not making public demands. However, making a public demand is a dominant strategy for each negotiator, and this leads to a suboptimal outcome. Escaping this prisoner's dilemma provides a rationale for secret negotiations. Testable hypotheses are derived from the nature of the commitments and agreements made in equilibrium.

**I**n many domestic and international bargaining situations, we often observe one or both negotiators making public statements in front of their constituents about the share of the benefits that they expect to obtain in the negotiations. For example, before the Copenhagen Summit of the European Union (EU) in December 2002, the Turkish government asked the EU to choose a date to start membership negotiations with Turkey. Anticipating that it was more likely that the EU would instead simply select a date to review whether Turkey had met membership conditions, the leader of Turkey's incumbent Justice and Development Party, Recep Tayyip Erdoğan, publicly announced that a review date was "not acceptable."<sup>1</sup>

Similarly, in the negotiations surrounding a peace deal in Northern Ireland in the mid-1990s, all of the parties involved made numerous public statements about their bargaining positions. For instance, in the lead-up to the negotiations that culminated in the "Good Friday Agreement" of April 1998, Prime Minister John Major of Britain declared that all of the Irish paramilitaries had to "decommission" their weapons before negotiations could begin. Similarly, the leader of the pro-union Ulster Unionist Party (UUP), David Trimble, publicly stated that there was "no question of negotiations without decommissioning." Meanwhile, Gerry Adams, head of the Irish Republican Army's (IRA) political wing, Sinn Féin, publicly announced that the

IRA's weapons would be decommissioned only after the conclusion of the negotiations.<sup>2</sup>

As a final example, since a tentative peace dialog began between India and Pakistan in January 2004, each side's government has repeatedly rebuked the other for presenting its bargaining position and intentions directly to the press rather than privately to the other government. President Pervez Musharraf of Pakistan has publicly stated that unless a final agreement is reached on the disputed region of Kashmir, talks on all other issues between the two sides, including trade, cross-border terrorism, and nuclear safeguards, would collapse. On the other hand, the Indian foreign minister publicly compared India's dispute with Pakistan to its traditional but recently declining tensions with China, implying that important progress on other issues could be made even if a final settlement on the border dispute remains elusive. The escalating public statements on both sides led the Pakistani foreign minister to call for a "rhetoric restraint regime."<sup>3</sup>

These examples pose a number of questions. First, what is the motivation behind making public statements like these, especially if backing down from them can carry domestic political costs? Second, why do the sides often make mutually incompatible public demands, since this means that at least one side's demands will go unfulfilled? Third, what is the incentive for each side to *restrain* itself from making public commitments, and will a "rhetoric restraint regime" ever be honored?

In this paper, we examine these issues by analyzing a game-theoretic bargaining model in which leaders can make public commitments prior to the bargaining, but backing down from these commitments is costly. That is, we assume that public statements generate potential "audience costs" for the leader (Fearon 1994, 1997; we discuss the possible sources of such costs later). Our analysis builds on Schelling's (1960, 28) intuition that,

Bahar Leventoglu is Assistant Professor, Department of Political Science, Stony Brook University, Stony Brook, NY 11794 (bahar.leventoglu@stonybrook.edu).

Ahmer Tarar is Assistant Professor, Department of Political Science, Texas A&M University, 4348 TAMU, College Station, TX 77843-4348 (ahmertarar@polisci.tamu.edu).

Our thanks to Huseyin Yildirim, the editor, three anonymous reviewers, and participants at seminars at New York University and Washington University in St. Louis, Mo (Political Economy Speaker Series), for helpful comments. The paper also benefited from comments received at the 2004 annual meetings of the American Political Science Association, the Midwest Political Science Association, and the Peace Science Society (International). The authors are responsible for any remaining shortcomings.

<sup>1</sup> Leyla Boulton and David Gardner, "An Islamist Seeking to be Europe's New Face," *Financial Times*, 7 December 2002.

<sup>2</sup> John Lloyd, "Ulster: Is Peace Now Worse than War?," *New Statesman*, 29 January 1999.

<sup>3</sup> Paul Watson, "India, Pakistan Schedule Talks," *Los Angeles Times*, 2 June 2004.

"When national representatives go to international negotiations knowing that there is a wide range of potential agreement within which the outcome will depend on bargaining, they seem often to create a bargaining position by public statements, statements calculated to arouse a public opinion that permits no concessions to be made." In particular, we explore how public commitments can be used to generate bargaining leverage in negotiations.

Ever since Schelling (1960), scholars in multiple disciplines have been interested in understanding the sources of strength in bargaining situations where the actors have common as well as conflicting interests. And for well over a decade now, many students of international relations have been intensely interested in moving beyond neorealism's treatment of the state as a unitary actor (Waltz 1979) and understanding the impact of domestic political factors on international relations.

Synthesizing these two trends, Putnam (1988) spurred a large amount of research on the effect on international bargaining of *exogenously* imposed domestic constraints on the executive, such as the requirement in many countries that major international agreements must be ratified by the legislature or by referendum (e.g., Iida 1993; Milner 1997; Mo 1994, 1995). In contrast, we investigate the much less-studied issue of how leaders can affect their bargaining position by *endogenously* imposing domestic "constraints" on themselves by making public statements that it would be costly to back down from (Pahre 1997).<sup>4</sup>

Our results speak to the old debate about whether the public nature of foreign policy decision making in democracies is a disadvantage or a benefit. Writers such as de Tocqueville ([1835] 1945) and Morgenthau (1956) have argued that effective diplomacy requires secrecy and freedom from domestic constraints. Our results indicate the conditions under which leaders as well as their citizens prefer negotiations to be held publicly or secretly. Contrary to the claims of de Tocqueville and Morgenthau, we show that publicity in negotiations can sometimes be an advantage.

The effects of audience costs have been explored in quite some detail in recent formal work on crisis bargaining, that is, bargaining in the shadow of war (e.g., Fearon 1994; Schultz 1999; Smith 1998). There has been much less work done on how audience costs can affect noncrisis bargaining, for example, the negotiation of trade agreements or treaties. We present such an analysis here.

## THE MODEL

The model is an extension of the following version of the Rubinstein (1982) bargaining model. Two players, labeled player 1 (a "she") and player 2 (a "he"), take turns making proposals to divide a pie of size 1. Negotiator 1 is chosen to make the first proposal with

probability  $0 \leq p \leq 1$ , and negotiator 2 makes the first proposal with probability  $1 - p$ , after which they alternate making proposals. Negotiators discount future payoffs with common discount factor  $0 < \delta < 1$ .

If player 1 is chosen to make the first proposal, let  $(x, 1 - x) \in \mathbb{R}^2$ , where  $0 \leq x \leq 1$  denote player 1's proposal. If player 2 accepts this proposal, then player 1 receives payoff  $x$ , player 2 obtains utility  $(1 - x)$ , and the game ends. If player 2 rejects the proposal, he makes a counterproposal in the next period, denoted by  $(1 - y, y) \in \mathbb{R}^2$ , where  $0 \leq y \leq 1$ . If player 1 accepts this proposal, player 1 obtains utility  $\delta(1 - y)$ , player 2 receives payoff  $\delta y$ , and the game ends. If player 1 rejects the proposal, she gets to make the next offer.

The game continues until one player accepts the other's proposal. In general, if an agreement  $z = (z_1, z_2)$  is reached in period  $t$  ( $t = 0, 1, 2, 3, \dots$ ), then player  $i$ 's payoff is  $\delta^t z_i$  ( $i = 1, 2$ ). If an agreement is never reached, both players receive utility 0. Rubinstein (1982) shows that there is a unique subgame perfect equilibrium (SPE) of this game in which the players always propose  $x = y = \frac{1}{1+\delta}$  for their own share and  $(1 - x) = (1 - y) = \frac{\delta}{1+\delta}$  for the other player's share, and in which the players reach an agreement in the first period of the game.

Here, we consider a variant of this model in which the two players (henceforth called negotiators) can make public commitments in front of their domestic constituents to obtaining some minimal share of the pie before the formal bargaining process begins. In the first move of the game, the two negotiators simultaneously announce their public commitments. Negotiator 1 publicly commits to receiving an amount of the pie at least equal to  $a$ , where  $0 \leq a \leq 1$ , and negotiator 2 commits to obtaining at least  $b$ , where  $0 \leq b \leq 1$ . If negotiator 1 receives at least  $a$  in the bargaining subgame, then her payoff is simply the share of the pie that she obtains (appropriately discounted by time). Otherwise, if she obtains less than  $a$ , then she pays a cost for backing down from her public commitment, and her overall payoff is the share of the pie minus the cost (appropriately discounted by time).

Let  $C_1(m, a)$  denote negotiator 1's cost for violating her public commitment when she commits to receiving at least  $a$  and actually receives  $m$ . Then we assume the following:

$$C_1(m, a) = \begin{cases} 0 & \text{if } m \geq a \\ \phi_1(a - m) & \text{otherwise, where } \phi_1 \geq 0. \end{cases}$$

Similarly, if negotiator 2 publicly commits to receiving at least  $b$  and actually receives  $n$ , then the cost he pays is

$$C_2(n, b) = \begin{cases} 0 & \text{if } n \geq b \\ \phi_2(b - n) & \text{otherwise, where } \phi_2 \geq 0. \end{cases}$$

The interpretation is that the cost increases linearly with the deficit between what the negotiator publicly commits to and what it actually receives: the greater the deficit, the greater the cost. The "cost coefficient"  $\phi$  measures how costly it is for the negotiator to violate

<sup>4</sup> Mo (1995) and Pahre (1997) examine how leaders may endogenously choose to impose ratification constraints on themselves.

a public commitment by a given amount: the higher  $\phi$  is, the more costly it is.

Note that in this model, in contrast to most previous formal models of audience costs, the magnitude of the audience cost is endogenous and depends on the negotiator's commitment level and the share of the pie it ends up accepting in equilibrium. The only part of the audience cost that is exogenous is  $\phi$ , and later we discuss how this parameter can vary by regime type and the leader's domestic political situation.

The timing of the game is as follows. First, the two negotiators simultaneously announce their public commitments  $a$  and  $b$ . Then nature chooses which negotiator gets to make the first proposal to divide the pie, with negotiator 1 chosen with probability  $0 \leq p \leq 1$  and negotiator 2 chosen with probability  $1 - p$ , after which they alternate. If they reach agreement on  $(z, 1 - z)$  in period  $t$  ( $t = 0, 1, 2, 3, \dots$ ), then player 1's payoff is  $\delta^t[z - C_1(z, a)]$  and player 2's utility is  $\delta^t[(1 - z) - C_2(1 - z, b)]$ . If the two negotiators never reach an agreement, both of them receive payoff 0.

In the economics literature, Muthoo (1992, 1996, 1999) also provides a formal analysis of the commitment tactic (so does Crawford 1982; however, he examines a very different type of problem and model). The primary way in which our work differs from his is that he uses the Nash bargaining solution (Nash 1950) to characterize the solution of the bargaining subgame, whereas we use an alternative-offers bargaining protocol and the subgame perfect equilibrium solution concept. Binmore (1987) shows that the unique subgame perfect equilibrium payoffs of the alternating-offers Rubinstein (1982) model converge to the Nash bargaining solution as the players' discount factor converges to one, and indeed our results converge to Muthoo's (1992) as the discount factor in our model converges to one. Thus, Muthoo's results emerge as a special case in our model, when the discount factor approaches one.

**Proposition 1.** For any  $\phi_1, \phi_2 \geq 0$ , the following is the unique stationary subgame-perfect equilibrium of this game: negotiator 1 makes the public commitment  $a^* = \frac{1+\phi_1}{1+\delta+\phi_1+\phi_2}$ , negotiator 2 makes the public commitment  $b^* = \frac{1+\phi_2}{1+\delta+\phi_1+\phi_2}$ , and when they do, in the bargaining subgame the negotiators use the following strategies:

- (a) Negotiator 1 always proposes  $(x^*, 1 - x^*) = (\frac{1+\phi_1}{1+\delta+\phi_1+\phi_2}, \frac{\delta+\phi_2}{1+\delta+\phi_1+\phi_2})$  and always accepts any proposal  $(1 - y, y)$  such that  $y \leq \frac{1+\phi_2}{1+\delta+\phi_1+\phi_2}$ .
- (b) Negotiator 2 always proposes  $(1 - y^*, y^*) = (\frac{\delta+\phi_1}{1+\delta+\phi_1+\phi_2}, \frac{1+\phi_2}{1+\delta+\phi_1+\phi_2})$  and always accepts any proposal  $(x, 1 - x)$  such that  $x \leq \frac{1+\phi_1}{1+\delta+\phi_1+\phi_2}$ .

Note that agreement is reached in the first period. If a player deviates from its equilibrium public commitment, then the strategies used in the bargaining subgame are specified in the proof in the appendix.<sup>5</sup>

We discuss this result in a number of parts.

<sup>5</sup> If  $\phi_i = 0$  ( $i = 1, 2$ ), then negotiator  $i$  can make any public commitment in equilibrium, as the commitment has no effect anyway.

## One-Sided Public Commitment

First consider the case where only one negotiator, say negotiator 1, pays a cost for backing down from a public commitment (i.e., suppose that  $\phi_1 > 0$  and  $\phi_2 = 0$ ). This might be the case, for instance, if country 1 is a democracy and country 2 is an autocracy. In this case, country 1's expected share of the pie is larger than what it would be if public commitments were not allowed (i.e., in the Rubinstein 1982 model), and country 2's is smaller.<sup>6</sup> In other words, being the only side to be able to make a costly public commitment provides bargaining leverage to that side. When a leader makes a public commitment that would be costly to back down from, that leader requires a larger share of the pie for it to be worthwhile to reach an agreement, and the other leader realizes this and hence compromises. Therefore, the public commitment tactic provides bargaining leverage.

Note that in equilibrium, the share of the pie that negotiator 1 proposes for herself when she makes a proposal is the same as her equilibrium public commitment (i.e.,  $x^* = a^*$ ). Hence, negotiator 1 does not pay an audience cost when she gets to make the first proposal (which negotiator 2 accepts). However, negotiator 2's proposal for negotiator 1 is less than negotiator 1's commitment level (i.e.,  $1 - y^* < a^*$ ). Therefore, negotiator 1 pays an audience cost when negotiator 2 gets to make the first proposal (which negotiator 1 accepts). Hence, negotiator 1's optimal commitment level in equilibrium is such that unless she makes the first proposal with certainty (i.e., unless  $p = 1$ ), she expects to pay an audience cost.

Because she expects to pay an audience cost, negotiator 1's expected payoff is a little less than her country's expected share of the pie. However, her payoff is still larger than it would be if public commitments were not allowed (i.e., in the Rubinstein 1982 model),<sup>7</sup> and hence the negotiator benefits from being the only side to generate costly public commitments. The other negotiator, on the other hand, is worse off.

## Two-Sided Public Commitment

When both sides face costs for backing down from public commitments (i.e., when  $\phi_1, \phi_2 > 0$ ), then whether public commitments are beneficial depends on the relative magnitudes of each side's audience cost rate,  $\phi_1$  and  $\phi_2$ .

In determining whether public commitments are beneficial to a side, there are two payoffs to consider. One is the country's share of the pie, which can be thought of as the welfare of the citizens of that country. The other is the negotiator's personal payoff, which is

<sup>6</sup> Negotiator 1's expected share of the pie is  $P_1(a^*, b^*) = p \cdot x^* + (1 - p)(1 - y^*)$ . It can easily be shown that when  $\phi_1 > 0$  and  $\phi_2 = 0$ ,  $x^* > \frac{1}{1+\delta}$  and  $1 - y^* > \frac{\delta}{1+\delta}$ .

<sup>7</sup> Negotiator 1's expected payoff is  $V_1(a^*, b^*) = p \cdot x^* + (1 - p)[(1 - y^*) - \phi_1[a^* - (1 - y^*)]] = p \cdot x^* + (1 - p)\delta x^*$ . It can easily be shown that when  $\phi_1 > 0$  and  $\phi_2 = 0$ ,  $x^* > \frac{1}{1+\delta}$  and  $\delta x^* > \frac{\delta}{1+\delta}$ .

the share of the pie minus the audience cost, if any, that is incurred.

It turns out that negotiator 1 benefits from public commitments if and only if she pays a significantly larger cost for violating a public commitment by a given amount than does negotiator 2 (i.e., if and only if  $\phi_1$  is sufficiently larger than  $\phi_2$ ). Why is this the case? First note that the equilibrium proposals for negotiator 1,  $x^*$  and  $1 - y^*$ , are both increasing in  $\phi_1$  and decreasing in  $\phi_2$ , which means that negotiator 1's expected share of the pie is also increasing in  $\phi_1$  and decreasing in  $\phi_2$ . Finally, note that negotiator 1's equilibrium public commitment  $a^*$  is also increasing in  $\phi_1$  and decreasing in  $\phi_2$ .

These results mean that as  $\phi_1$  increases (or  $\phi_2$  decreases), negotiator 1 is demanding a bigger share of the pie and getting more. The net result is that her expected payoff is increasing in  $\phi_1$  and decreasing in  $\phi_2$ . The more costly it is for a negotiator to violate a public commitment by a given amount (and the less costly it is for the other side), the greater its equilibrium public commitment, its share of the pie, as well as its personal payoff.<sup>8</sup>

It turns out, then, that whether a negotiator benefits from public commitments depends on the relative values of  $\phi_1$  and  $\phi_2$ . In particular, negotiator 1 benefits from public commitments (relative to the Rubinstein 1982 model in which public commitments are not allowed) if and only if  $\phi_1$  is sufficiently larger than  $\phi_2$  (in particular, if and only if  $\phi_1 > \frac{\phi_2}{\delta}$ ). Similarly, negotiator 2 benefits from public commitments if and only if  $\phi_2$  is sufficiently larger than  $\phi_1$  (in particular, if and only if  $\phi_2 > \frac{\phi_1}{\delta}$ ). When  $\phi_1$  and  $\phi_2$  are close to each other, *both* sides are worse off than they would be without public commitments.

We normally think that democratic leaders pay significantly greater costs for violating public commitments than do autocratic leaders who are less accountable to the public (e.g., Fearon 1994 uses this as a working assumption; also see Schelling 1960, 28); that is, a democratic leader has a significantly greater  $\phi$  than does an autocratic leader. Thus, a prediction of the model is that democratic leaders can and will use public commitments to obtain bargaining leverage when negotiating with autocratic leaders. On the other hand, the cost of losing power for autocrats is often higher than for democratic leaders, including the possibility of imprisonment or execution, among others (Gowa 1995). Therefore, this assumption does not always have to hold.<sup>9</sup>

Even if two democratic leaders are negotiating with each other, they may differ quite a bit in how costly it is to violate a public commitment by a given amount. For example, it seems that violating a public commitment would be especially costly just prior to elections,

because elections provide a particularly convenient method for voters to punish their leader for violating a public commitment. This would especially be the case if the leader is also politically vulnerable domestically, for example, if it is facing a weak economy or other domestic problems. We might call this type of leader, which (we presume) has a very high  $\phi$  because it is facing elections *and* is politically insecure, a high audience cost leader.

A leader who is domestically secure and not facing elections would seem to face the lowest cost for violating a public commitment, and we might call this a low audience cost leader. Leaders who are facing elections but are politically secure, as well as leaders who are politically vulnerable but are not facing elections, would seem to have an intermediate cost for violating a public commitment, and we might call these medium audience cost leaders. The model predicts that a high audience cost leader would be able to use public commitments to gain bargaining leverage when negotiating with a low audience cost leader, and possibly with medium audience cost leaders as well, depending on the difference in their audience cost coefficients. Similarly, medium audience cost leaders may have bargaining leverage when negotiating with a low audience cost leader.

An autocratic leader who is domestically vulnerable may have bargaining leverage when negotiating with a low audience cost democratic leader. And all types of leaders who face positive audience costs can generate bargaining leverage when negotiating with entities that do not, such as when developing countries are negotiating with international institutions such as the International Monetary Fund (IMF) for the terms of financial assistance.

## A Prisoner's Dilemma and a Rationale for Secret Negotiations

We have seen that when only one side can generate costly public commitments or one side pays a significantly greater cost for violating a public commitment by a given amount than does the other, then the former negotiator benefits from public commitments and the latter is worse off. On the other hand, if  $\phi_1$  and  $\phi_2$  are both positive and close to each other (in particular, if  $\delta\phi_2 < \phi_1 < \frac{\phi_2}{\delta}$ ), then *both* negotiators are worse off with public commitments than without.

To understand why this is the case, consider the situation where  $\phi_1 = \phi_2 = \phi > 0$  and  $p = 1/2$  (i.e., each side has an equal chance of being chosen to make the first proposal, so there is no first-mover advantage in expectation). Then each side makes the same public commitment  $a^* = b^* = \frac{1+\phi}{1+\delta+2\phi}$ , which is greater than  $1/2$ , but each side only expects to receive  $1/2$  in the bargaining subgame. That is, each side expects to obtain merely the same amount of the pie that it would if public commitments were not allowed, but also pays an audience cost with positive probability (if the other side is chosen to make the first proposal). Hence, both sides would be better off if neither made a public

<sup>8</sup> Muthoo (1992) also finds that a negotiator's payoff is increasing in its cost coefficient.

<sup>9</sup> Another way of saying this is that democratic leaders face a greater *likelihood* of losing office for backing down from a public commitment, but the *payoff* for this outcome can be significantly worse for autocrats.



**TABLE 1. The Prisoner's Dilemma Induced by the Ability of Both Sides to Make Costly Public Commitments**

		Negotiator 2	
		No Commitment	Commitment
Negotiator 1	No Commitment	$\frac{p(1-\delta)+\delta}{1+\delta}, \frac{1-p(1-\delta)}{1+\delta}$	$\frac{p(1-\delta)+\delta}{1+\delta+\phi_2}, \frac{(1+\phi_2)[1-p(1-\delta)]}{1+\delta+\phi_2}$
	Commitment	$\frac{(1+\phi_1)[p(1-\delta)+\delta]}{1+\delta+\phi_1}, \frac{1-p(1-\delta)}{1+\delta+\phi_1}$	$\frac{(1+\phi_1)[p(1-\delta)+\delta]}{1+\delta+\phi_1+\phi_2}, \frac{(1+\phi_2)[1-p(1-\delta)]}{1+\delta+\phi_1+\phi_2}$

commitment.<sup>10</sup> The same general result holds whenever  $\phi_1$  and  $\phi_2$  are close to each other. We would expect this to be the case, for instance, when two leaders facing similar domestic political conditions negotiate with each other.

This suggests that prior to entering into negotiations, two farsighted leaders facing fairly similar costs for violating public commitments would make an agreement to refrain from making public commitments. This resembles the "rhetoric restraint regime" proposed by the Pakistani foreign minister, as discussed earlier. However, the problem turns out not to be so simple, because the ability of both sides to make public commitments actually creates a prisoner's dilemma in which each side has a dominant strategy of making a public commitment.

Table 1 shows the strategies and the resulting payoffs. Each side's most preferred outcome is where it makes a public commitment but the other side does not. Each side's least preferred outcome is where it does not make a public commitment but the other side does. And when  $\phi_1$  and  $\phi_2$  are sufficiently close to each other (in particular, when  $\delta\phi_2 < \phi_1 < \frac{\phi_2}{\delta}$ ), then each side prefers the outcome where neither makes a public commitment to the outcome where both make public commitments.

This preference ordering induces the familiar prisoner's dilemma in which each side's dominant strategy is to make a public commitment (in the traditional parlance of the prisoner's dilemma, to "defect"). If you believe that the other side is not going to make a public commitment, you want to make one in order to obtain the bargaining leverage of the one-sided case; and if you believe that the other side is going to make a public commitment, you also want to make one in order to mitigate the bargaining leverage that the other side will otherwise have over you. Thus, no matter what you believe that the other side is going to do, you are best off making a public commitment. Each side's dominant strategy leads to the suboptimal outcome where both make public commitments, an outcome which is Pareto-dominated by both not making public commitments.

Therefore, the model illustrates how difficult it is for two leaders facing fairly similar costs for violating a public commitment by a given amount to refrain from

making public commitments and winding up in a suboptimal outcome. This suggests that any nonbinding "rhetoric restraint regime" is unlikely to work. And indeed, just prior to a meeting between the two countries' foreign ministers in early September 2004, an Indian foreign ministry spokesman stated that, "There is considerable disappointment here today at the unifocal statement made by the Pakistan foreign minister about India-Pakistan relations. . . . This is not in consonance with the spirit in which we have conducted the composite dialogue so far. *It also violates Pakistan's own call for a rhetoric restraint regime*" (emphasis added).<sup>11</sup>

However, the possibility of keeping the negotiations secret provides a solution to this problem. Although neither side will abide by an unenforceable agreement not to make a public commitment because the dominant strategy is to make a public commitment, if the negotiations are being conducted secretly without the public's knowledge, then there is nothing to publicly commit to; hence, the suboptimal outcome can be avoided. Conducting the negotiations secretly provides a mechanism for both sides to avoid making public commitments and winding up in the suboptimal outcome. Moreover, detecting violations of such an agreement is relatively easy, making countries more willing to rely on them (e.g., Keohane 1984). Hence, our model provides a new rationale for secret negotiations.

For example, the negotiations that led to the 1993 Oslo Accords between the Israelis and the Palestinians were conducted secretly and only made public once an agreement had been reached. Public talks sponsored by the United States were occurring at the same time in Washington with a different Palestinian negotiating team. The Washington talks were publicly known, and the Palestinian team there was making large demands regarding settlements and Jerusalem that the Israeli team found unacceptable (Perlmutter 1995). An agreement was only able to be reached in the secret negotiations being held in Oslo.

Subsequent negotiations between the two sides have taken place in the public eye and have been much more difficult to negotiate, up to the point that in recent years the negotiation process has almost

<sup>10</sup> They are better off ex ante as well as ex post, that is, before nature chooses the first proposer as well as afterwards.

<sup>11</sup> Denyer, Simon, "Signs of Discord as Indo-Pakistan Ministers Meet," *Reuters*, <<http://www.reuters.com/newsArticle.jhtml?type=worldNews&storyID=6152478&section=news>> (September 5, 2004).

completely come to a halt. These subsequent negotiations have occurred amidst public posturing on both sides. For example, regarding the final status of Jerusalem, which the Oslo Accords left for future negotiations, the late Palestinian leader Yasser Arafat repeatedly made public statements promising that Jerusalem would become the capital of a Palestinian state, whereas Yitzhak Rabin and subsequent Israeli prime ministers have made public promises that Jerusalem would remain the undivided capital of Israel (Perlmutter 1995).

Arafat also made numerous statements promising to secure a right of return for Palestinian refugees to their former homes in Israel, whereas all Israeli prime ministers have publicly declared that that is not an option. Makovsky (2001) writes:

The process also allowed each side to make contrary claims at home. . . . Israeli leaders were able to continually promise their constituents what they wanted—including a united Jerusalem under Israeli sovereignty—while Arafat could promise his people what they wanted—including the right of return for all Palestinians to long-abandoned homes inside Israel. Arafat sold Oslo to his public by telling them it guaranteed a return to the 1967 lines and entailed no compromises. He led his people to believe that they would get 100 percent of the land they wanted.

When Arafat was not offered all of what he had promised to his people at the 2000 Camp David talks, in particular a right of return for the Palestinian refugees, the talks ended without an agreement, and the negotiation process ground to almost a complete halt soon afterwards when the second Intifadah began and Ariel Sharon was elected prime minister of Israel. Our model, which does not incorporate third-party actors such as extremists who can scuttle an agreement by diminishing trust between the two sides (see Kydd and Walter 2002), does not predict that an agreement will fail to be reached—it does predict, however, that the two sides will make mutually incompatible public demands and that the two leaders will have less of an incentive to reach an agreement than if public commitments were not allowed.

Indeed, it turns out in equilibrium that  $a^* + b^* > 1$ : the two sides make mutually incompatible public demands; that is, the sum of their commitments exceeds the amount of pie that is available to be divided.<sup>12</sup> Hence, at least one side gets less than what it publicly committed to (*ex post*, exactly one side gets less in equilibrium, but *ex ante* both sides expect to get less whenever  $0 < p < 1$ ). A consequence of this is that the ability to make public commitments leads to an inefficiency in the bargaining outcome for the negotiators, because the negotiator that does not get to make the first proposal pays an audience cost in equilibrium. In addition to the Israeli–Palestinian case just mentioned, mutually incompatible public commitments were also made in the three cases discussed in the introduction.

As another example, Israeli and Egyptian leaders made mutually incompatible public commitments in the course of their tentative peace overtures to each other in the late 1970s. In a speech to the Israeli Knesset during an historic visit to Israel in November 1977, Egyptian president Anwar el Sadat stated that Egypt would only make peace with Israel if all Arab territories captured in the 1967 Six Day War were returned. During a visit to the Egyptian city of Ismailiyya the following month, Israeli prime minister Menachem Begin proposed that “autonomy” would be granted to Palestinians in the West Bank and Gaza Strip, but those territories would remain under Israeli sovereignty. Sadat rejected this as unacceptable, and the tentative peace process came close to a halt.

The Carter administration then intervened and after many months of discussion with both sides, Sadat and Begin agreed to meet with Carter at Camp David in September 1978. Partly to isolate each side from domestic pressures (and perhaps to make it easier for each side to restrain itself from making public commitments), Carter insisted that no reporters and television cameras be allowed during the course of the negotiations. Unlike the Oslo negotiations, the outside world *was* aware that negotiations were taking place—however, the negotiators were secluded from the press until the negotiations concluded after 13 days with an agreement that would eventually become a peace treaty between Egypt and Israel (Telhami 1990). In this case, a third party (the United States) was able to enforce a ban on public statements by hosting the negotiations under controlled conditions, which suggests another possible solution to the prisoner’s dilemma.

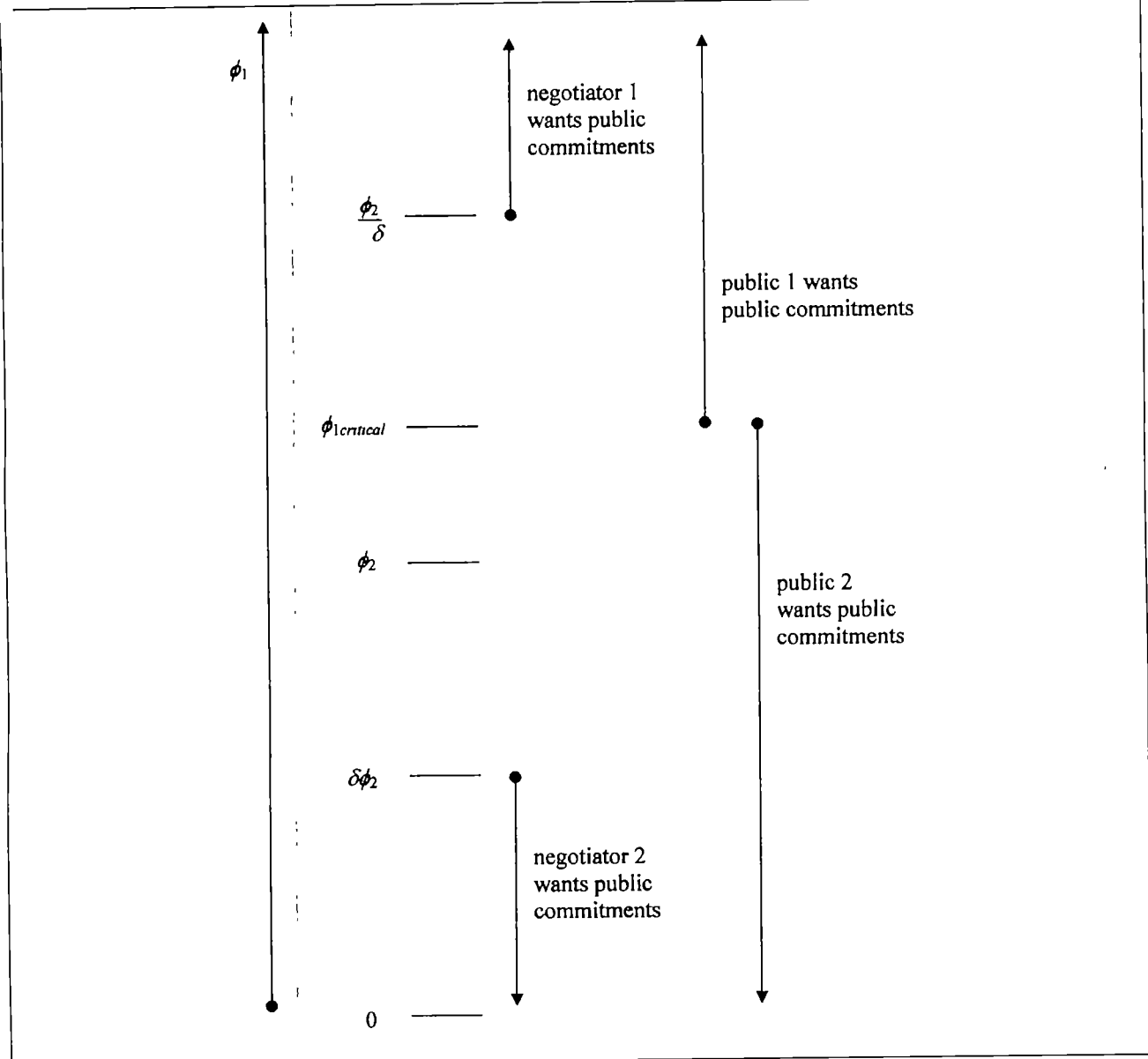
## A Principal–Agent Problem

So far, we have been examining the payoffs of the negotiators and their incentives to conduct the negotiations secretly or publicly. However, examining the payoffs of the citizens shows that there is a type of principal–agent problem that can arise from the ability to make costly public commitments.

We saw that negotiator 1 prefers public commitments to no commitments if and only if she pays a significantly greater cost for violating a public commitment by a given amount than does negotiator 2, that is, if and only if her cost coefficient  $\phi_1$  is significantly larger than  $\phi_2$ . Recall that the negotiator’s payoff is the share of the pie minus the audience cost, if any, that is incurred. Because they do not incur audience costs, only the leader does, the payoff of the *citizens* of a country can be thought of as simply that country’s share of the pie. Recall that country 1’s share of the pie (as well as negotiator 1’s personal payoff) is increasing in  $\phi_1$  and decreasing in  $\phi_2$ . Because of this, country 1’s share of the pie with public commitments is larger than what it would be without public commitments if and only if  $\phi_1$  is sufficiently large relative to  $\phi_2$ . However, it turns out that the threshold that  $\phi_1$  has to exceed is

<sup>12</sup> Muthoo (1992) finds that the sum of the commitments is exactly 1. This is a special case of our results, since  $a^* + b^* \rightarrow 1$  (from above) as  $\delta \rightarrow 1$ .

**FIGURE 1. Range of Values of  $\phi_1$  Relative to  $\phi_2$  for Which Negotiators 1 and 2 and Their Publics Benefit from Public Commitments**



not as large for the country's share of the pie as it is for the negotiator's payoff.

This is illustrated in Figure 1. This figure shows the range of values of  $\phi_1$  relative to  $\phi_2$  for which the leaders of countries 1 and 2 as well as their citizens want public commitments rather than no commitments. Leader 1 wants public commitments, if and only if  $\phi_1$  is sufficiently larger than  $\phi_2$  (in particular, if and only if  $\phi_1 > \frac{\phi_2}{\delta}$ ), and leader 2 wants public commitments if and only if  $\phi_1$  is sufficiently smaller than  $\phi_2$  (in particular, if and only if  $\phi_1 < \delta\phi_2$ ). Neither leader wants public commitments if  $\phi_1$  and  $\phi_2$  are close to each other (in particular, if  $\delta\phi_2 < \phi_1 < \frac{\phi_2}{\delta}$ ). The citizens of country 1 want public commitments if and only if  $\phi_1 > \phi_{1critical}$  (where  $\phi_{1critical} = \frac{(\delta+p-\phi)\phi_2}{1+p\delta-p}$ ), and the citizens of country 2 want public commitments if and only if  $\phi_1 < \phi_{1critical}$ . Of main importance, as shown in the figure, is that

the threshold of the citizens,  $\phi_{1critical}$ , lies between the thresholds of the negotiators.<sup>13</sup>

The basic intuition behind this is that because they do not pay an audience cost, only the leader does, the citizens of country 1 have a lower threshold for  $\phi_1$  above which they prefer public commitments to no commitments than does their leader. Hence, they want the negotiations to be held publicly under some conditions in which their leader wants them to be held secretly (namely, when  $\phi_1$  is somewhat large but not too large). The same is true for the citizens of country 2.

<sup>13</sup> Note that  $\phi_{1critical}$  is an increasing function of  $\phi_2$  and that  $\phi_{1critical} \in (\delta\phi_2, \frac{\phi_2}{\delta})$  for  $p \in (0, 1)$ . Also note that  $\phi_{1critical} \rightarrow \frac{\phi_2}{\delta}$  (from below) as  $p \rightarrow 1$ , and  $\phi_{1critical} \rightarrow \delta\phi_2$  (from above) as  $p \rightarrow 0$ . When  $p = 1/2$ ,  $\phi_{1critical} = \phi_2$ .

Therefore, as seen in the figure, for *all* values of  $\phi_1$  relative to  $\phi_2$  (except the knife-edge case where  $\phi_1 = \phi_{1critical}$ ), it is the case that at least one of the four “actors” strictly wants public commitments. Even if  $\phi_1$  and  $\phi_2$  are close enough to each other that neither leader wants public commitments, one (and only one) of their publics wants public commitments.

This result has a number of implications. It can be seen from Figure 1 that our model predicts that (under complete information) it is *never* the case that both executives want public negotiations—if one side is benefiting, the other is worse off. However, in the real world we often observe public negotiations occurring. One answer to this puzzle is that public negotiations are the “normal” way of negotiating and that secret negotiations require the active assent of both parties. If one side objects to secret negotiations, the negotiations will be held publicly (and indeed, in our model in which no gains are made if the two sides do not negotiate, even the negotiator that does not want to negotiate publicly gains more from negotiating publicly than from not negotiating at all).

Another answer is incomplete information: if the two sides are uncertain of the other side’s audience cost rate  $\phi$ , *each* might believe that it will benefit from public commitments. An incomplete information extension of this model would be worthwhile for future research.

Finally, Figure 1 suggests a domestic politics-based explanation for why negotiations might be held publicly even when both negotiators want them held secretly. As seen from the figure, even when both leaders have an incentive to keep the negotiations secret (i.e., when  $\phi_1$  and  $\phi_2$  are close to each other), one side’s public wants the negotiations to be held publicly. Although we do not explicitly model this, if this public can impose sufficient ex post costs on their leader for negotiating secretly, the leader will want to negotiate publicly even though its own preference (absent that cost) is to negotiate secretly. The citizens thus “force” their leader to go public and incur audience costs in order to bring the citizens net benefits. This suggests that the secret negotiations mechanism will be hard for one negotiator to implement, if it anticipates that an ex post cost will be imposed by its citizens for negotiating secretly.

This also provides an explanation for the conventional wisdom that democratic publics dislike secret negotiations. One common explanation for this is that the people are suspicious that their leader is secretly “giving away the store.” For example, such an interpretation could be applied to Israeli prime minister Ehud Barak’s relatively large concessions to Palestinian leader Yasser Arafat during the 2000 Camp David negotiations, and was also part of the basis for Woodrow Wilson’s call for “open covenants, openly arrived at” (Jordan, Taylor, and Mazarr 1999, 54). However, the “giving away the store” explanation assumes that the public and the leader have (or may have) quite different preferences. In our model, the two have the same basic preference: they both want to obtain as large a share of the pie as possible for their country. However, there exist circumstances in which the leader

wants the negotiations to be held secretly because it will otherwise incur audience costs which are greater than its side’s increase in the share of the pie, but the public knows that the leader will obtain a larger share of the pie with public negotiations and hence does not want the leader to hold them secretly. The model thus provides an explanation for the conventional wisdom that democratic publics often dislike secret negotiations, without assuming that the leader has (or might have) different preferences from the majority of the public.

Finally, note that these results speak to the old question of whether the public nature of foreign policy decision making in democracies is a disadvantage or a benefit. Writers such as de Tocqueville ([1835] 1945) and Morgenthau (1956) have argued that effective diplomacy requires secrecy and freedom from domestic constraints. However, our results indicate that under some conditions a negotiator as well as its citizens benefit from public negotiations. Contrary to the claims of de Tocqueville and Morgenthau, publicity in negotiations can sometimes be an advantage.

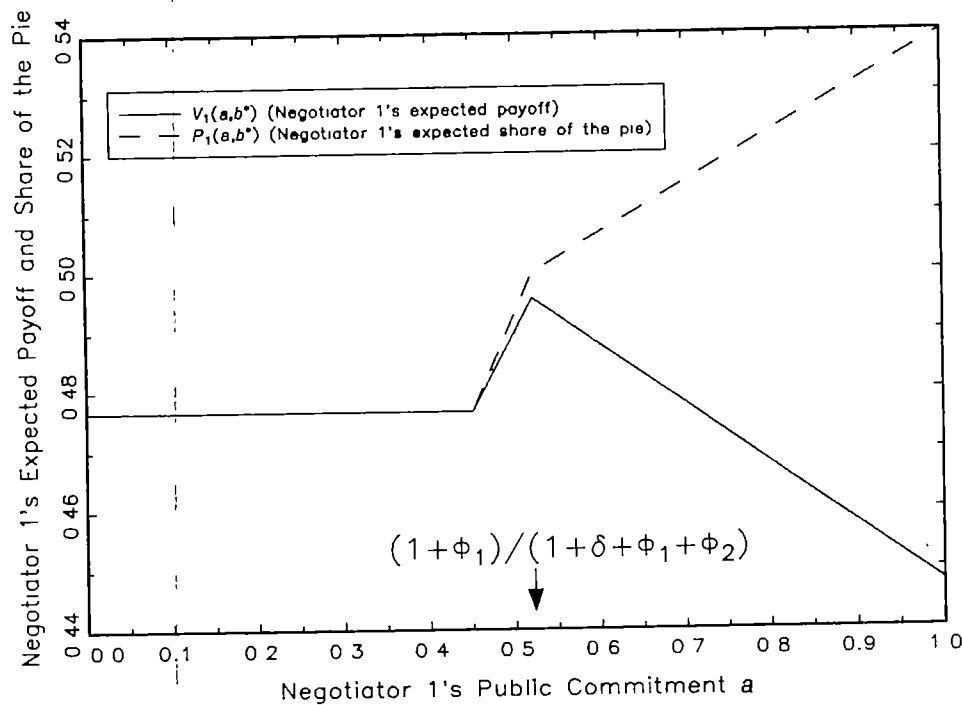
## Another Principal-Agent Problem

It turns out that there is another type of principal-agent problem that arises from the ability to make costly public commitments: namely, the negotiator does not make as large a public commitment as its citizens would like.

This is seen in Figure 2, which shows negotiator 1’s expected share of the pie  $P_1(a, b^*)$  and expected payoff  $V_1(a, b^*)$  (share of the pie minus the audience cost, if any, that is incurred) as a function of her public commitment  $a$  as  $a$  ranges from 0 to 1, when negotiator 2 is choosing his equilibrium commitment level  $b^*$ .

As seen from the figure, when negotiator 1 makes a very low public commitment, then the commitment is too low to have any effect on the bargaining subgame. Extremely low public commitments have no effect, because the share of the pie that goes to the negotiator if she did not make a commitment is enough to satisfy a low commitment, and so it is as if no commitment were made.

On the other hand, when negotiator 1’s public commitment  $a$  gets in the medium range, then her expected share of the pie starts increasing in  $a$  because the higher her public demand, the bigger are negotiator 1 and 2’s equilibrium proposals for negotiator 1,  $x$  and  $1 - y$ , respectively. In this region, negotiator 1’s expected utility is slightly lower than her expected share of the pie, because when negotiator 2 gets to make the first proposal, he offers negotiator 1 less than her public commitment ( $1 - y < a$ ), and so negotiator 1 pays an audience cost. However, the difference between negotiator 1’s share of the pie and her personal payoff is only slight, because when negotiator 1 is chosen to make the first proposal, her proposal for herself is larger than her public commitment ( $x > a$ ); hence, she does not pay an audience cost in this case.

FIGURE 2. Negotiator 1's Expected Payoff  $V_1(a, b^*)$  and Expected Share of the Pie  $P_1(a, b^*)$ 

Once negotiator 1's public commitment  $a$  gets too large, however, her expected payoff starts decreasing in  $a$ . She is still getting bigger and bigger offers ( $x$  and  $1 - y$ ); hence, her share of the pie is still increasing in  $a$ . However, these offers are now increasing at a smaller rate than before. More importantly, her public commitment  $a$  is now high enough that even negotiator 1's own proposal for herself is less than her public commitment:  $x < a$  in addition to  $1 - y < a$ . Therefore, although she is still getting bigger offers, she is now *always* paying a cost for violating her public commitment, and the net result is that her expected payoff is decreasing in  $a$ .

Therefore, as seen in the figure, negotiator 1's expected payoff is maximized at  $a^* = \frac{1+\phi_1}{1+\delta+\phi_1+\phi_2}$ , and in equilibrium, this is the public commitment that she makes.<sup>14</sup> One implication of Figure 2 is that the negotiator does not make the commitment that maximizes the welfare of her *citizens*. Because the share of the pie is always increasing in the commitment level, the citizens want their negotiator to demand the entire pie. However, the negotiator chooses not to do this, because the cost she would pay for getting less than her public commitment makes it not worthwhile.

This illustrates an interesting point. It is the credibility that the leader will be punished for backing down from a public commitment that allows the leader to use a public commitment to extract a bargaining concession from the other side, a concession that benefits

both the leader and the public; however, it is this very credibility that also ensures that the leader will not use the commitment tactic to the public's maximum advantage. The public's ability to impose costs on their leader is to their benefit; however, it also ensures that the benefit will not be all that it can be. The ability to make a public commitment generates a principal-agent situation in which the agent brings benefits to the principal (and to itself), but the agent's own interests limit the extent of the principal's benefits.

### Equilibrium Public Commitments and Offers

The final interesting result to note from Proposition 1 is that in equilibrium  $x^* = a^*$  and  $y^* = b^*$ . That is, each negotiator's proposal for itself is the same as its public commitment. Hence, a negotiator never pays an audience cost in its own proposal (which is accepted by the other negotiator). However,  $1 - y^* < a^*$  and  $1 - x^* < b^*$ . That is, a negotiator's share of the pie when the *other* side makes a proposal is less than its public commitment; hence, each negotiator pays an audience cost in the other side's proposal (which it accepts).<sup>15</sup>

Hence, if the two negotiators probabilistically decide who gets to make the first proposal (or if they are uncertain about who will get to make the first proposal), then *each* side's optimal commitment in equilibrium is such that it expects to pay an audience cost. However,

<sup>14</sup> Note that as  $\delta \rightarrow 1$ , the equilibrium public commitments in Proposition 1 converge to those in Muthoo's (1992, 383) Proposition 1.

<sup>15</sup> In Muthoo's (1992) results, each side's proposal offers each side exactly its public commitment, and so audience costs are never paid. In our results, this is a special case since  $1 - y^* \rightarrow a^*$  and  $1 - x^* \rightarrow b^*$  (from below) as  $\delta \rightarrow 1$ .

because the equilibrium does not depend on the value of  $p$ , even if one negotiator knows for certain that it will not make the first offer (i.e., even if  $p = 0$  or  $p = 1$ ), it chooses a public commitment high enough that it knows that it will pay an audience cost.

The intuition behind this is as follows: by making a higher public commitment, a negotiator makes it more likely that it falls in the range where it pays an audience cost, as well as increases the magnitude of that cost (when it is in the range where an audience cost is paid). However, it also increases the share of the pie that it obtains. In equilibrium, the optimal tradeoff is that the negotiator chooses to pay a limited audience cost in order to obtain a larger share of the pie.

This provides a rationale for why leaders typically make greater public demands than they expect to actually achieve (in fact, the model predicts that the optimal demand is such that the negotiator might obtain as much, but never more). Although the negotiator expects to pay a cost for doing so, the increased share of the pie that the commitment leads to more than compensates for this. The model explains why leaders publicly demand a lot, but not as much as the citizens would like.

## CONCLUSION

In this paper, we analyze a formal model to help explain why negotiators often publicly commit themselves to obtaining certain benefits prior to entering into negotiations. And although most of our examples have been drawn from international politics, we believe that this bargaining tactic is also often used in domestic negotiations. For example, in the early 1950s, there was an attempt in the U.S. Senate to pass a constitutional amendment that would have put severe limits on the president's ability to negotiate executive agreements with other countries that do not require congressional approval. The Eisenhower administration first used quiet means to try to sink the Bricker amendment (named after its sponsor, Senator John W. Bricker (R-Ohio)), for example, by supporting an alternative, weaker version of the amendment. But when it became clear that the Bricker amendment was likely to pass on the Senate floor, the administration escalated to open confrontation, including placing an open letter in the *Congressional Record* stating that the president was "unalterably" opposed to the amendment (Martin 2000, 77). Ultimately, the amendment was defeated. Under the reasonable supposition that because of far greater media coverage the president can generate much greater audience costs than individual senators can, this outcome is consistent with the predictions of our model.

There has been much discussion in the crisis bargaining literature (bargaining in the shadow of war) on the microfoundations of audience costs (e.g., Schultz 1999; Smith 1998). At least four possibilities have been discussed in the literature, and our results suggest an additional one. First, Fearon (1994) argues that a leader that backs down from a public commitment may pay a do-

mestic cost (e.g., is less likely to be reelected) because the domestic audience has perceived that the leader has violated the "national honor." Second, he and Smith suggest that a leader that makes a public commitment and then has to back down from it may be perceived by the domestic public to be *incompetent*, hence, be less likely to be reelected. Third, Sartori (2002) argues that a leader caught bluffing may pay an *international* audience cost because that leader's rhetoric is less likely to be considered credible by leaders of other countries in the future: the cost is due to loss of future international credibility. Finally, Guisinger and Smith (2002) point out that this international audience cost can also lead to a domestic audience cost: if the rhetoric of a leader caught bluffing is less likely to be believed by other leaders in the future, and this leads to welfare losses for the nation because its diplomacy lacks effectiveness, this may be reason for the public to depose that leader and insert a new one with a fresh reputation.

We believe that all of these arguments, especially the last one, have merit. However, our own analysis suggests an additional rationale for audience costs. In our model, the ability to generate audience costs provides bargaining benefits to a negotiator, benefits that accrue to the public as well. Therefore, in a repeated negotiations framework in which a country is repeatedly negotiating international agreements, if the public's strategy is to punish a leader (perhaps electorally) who violates a public commitment, then this strategy allows their leader to generate audience costs and to secure bargaining benefits for them. On the other hand, if their strategy is to *not* punish their leader for violating a public commitment, then no extra bargaining leverage is obtained. Hence, voters in a democracy have an incentive to punish their leader for violating a public commitment not because of any vindictive or "national honor" related reasons, but simply because such a strategy provides them with a stream of bargaining benefits over the long run.<sup>16</sup>

Our results have potentially important implications for the literature on signaling in international crises. Previous analyses of audience costs focus on a crisis bargaining setting in which two countries are in a dispute over an indivisible good and each is uncertain of the other's resolve for going to war (e.g., Fearon 1994, 1997; Schultz 1999). In this setting, it is argued, leaders (especially of democracies) can credibly convey their resolve by making public threats that generate potential audience costs. This literature concludes that audience costs, by allowing for credible information transmission in an incomplete information setting, generally has a beneficial effect: it reduces the frequency of

<sup>16</sup> Schultz (1999, 237) makes a related argument in the context of crisis bargaining—namely, that higher audience costs make it more likely that a state will prevail in international crises (Fearon 1994), and this is a rationale for the citizens to impose audience costs. Our argument focuses on bargaining benefits, and hence is related, but also quite different. In a technical supplement to this article, which is available from the authors' web sites or on request from the authors, we analyze a formal repeated negotiations model and derive the conditions under which the citizens will rationally choose to impose audience costs on their leader.

suboptimal wars due to incomplete information (e.g., Fearon 1995).

However, our model, which considers a divisible good and allows for genuine bargaining, shows that audience costs can also be used as an instrumental source of bargaining leverage, even in a complete information setting in which audience costs have no signaling value. Moreover, the *mutual* use of audience costs can lead to a suboptimal outcome for both sides. Although ours is not a crisis bargaining model, its results suggest that in crisis bargaining over a divisible good, the use of audience costs as an instrumental source of bargaining leverage may lead to suboptimal outcomes, perhaps even war. If this is the case, we would have to reconsider the traditional beneficial view of audience costs that we have due to their signaling value. We leave this important examination of the effects of audience costs in crisis bargaining for future research.

Another important extension of this paper would be allowing for incomplete information about the other side's audience cost coefficient  $\phi$ . With uncertainty like this, the two sides might end up making commitments that are jointly so large that there no longer exist agreements that both negotiators prefer to the status quo; hence, negotiations would break down (e.g., Muthoo 1999; this is a persuasive explanation of the failure of the 2000 Israeli-Palestinian Camp David negotiations, prior to which Arafat committed to a right of return for the Palestinian refugees, a concession which was not granted by Barak). Some of our findings might be modified under incomplete information. For example, we find that the public wants its leader to make as large a public commitment as possible, as this secures the largest share of the pie. With incomplete information, however, this incentive may no longer exist, as a larger public commitment probably would lead to a larger probability of negotiation failure. This also suggests, however, that leaders might modify their public commitments in response to the other side's commitment. We leave these important extensions of the model for future research.

## APPENDIX

### Proof of Proposition 1

Here, we prove that the strategies described in Proposition 1 comprise a subgame perfect equilibrium (SPE). In a technical supplement to this article, which is available from the authors' web sites or on request from the authors, we provide a proof that this is the unique stationary SPE.

We conjecture that there exists a SPE of the game in which negotiator 1 always proposes some  $(x, 1-x)$  and negotiator 2 always proposes some  $(1-y, y)$ , and these proposals are accepted. Also, the equilibrium proposals and commitments satisfy  $x, 1-y \leq a$  and  $y, 1-x \leq b$ . That is, each side's proposal offers each side no more than its public commitment. (In equilibrium, it turns out that  $x = a, 1-y < a, y = b$ , and  $1-x < b$ . But we adopt a more general approach in deriving the equilibrium.)

Our approach is to first determine for which values of the commitments  $a$  and  $b$  there exists a SPE of the bargaining subgame in which such proposals are made. We then identify

an equilibrium level of commitments  $a^*$  and  $b^*$  such that each player is strictly worse off by choosing a different commitment level, when the other player is choosing its equilibrium commitment level.

In the conjectured SPE of the bargaining subgame, negotiator 1 proposes  $(x, 1-x)$  and negotiator 2 accepts it. Negotiator 2 proposes  $(1-y, y)$  and negotiator 1 accepts it. For negotiator 2 to accept 1's proposal, negotiator 2's overall payoff for accepting it should be (at least) equal to his overall payoff if he rejects negotiator 1's proposal, makes a counter proposal himself, and negotiator 1 accepts it. Moreover, for negotiator 1 to accept negotiator 2's proposal, negotiator 1's overall payoff if she accepts negotiator 2's proposal should be (at least) equal to her overall payoff if she rejects negotiator 2's proposal, makes a counter proposal herself, and negotiator 2 accepts it. That is,

$$(1-x) - \phi_2(b - (1-x)) = \delta[y - \phi_2(b-y)]$$

$$(1-y) - \phi_1(a - (1-y)) = \delta[x - \phi_1(a-x)].$$

Solving this pair of simultaneous equations for  $x$  and  $y$ , we obtain

$$x = \frac{1}{1+\delta} + \frac{\delta(1+\phi_2)\phi_1 a - (1+\phi_1)\phi_2 b}{(1+\phi_1)(1+\phi_2)(1+\delta)}$$

$$y = \frac{1}{1+\delta} + \frac{\delta(1+\phi_1)\phi_2 b - (1+\phi_2)\phi_1 a}{(1+\phi_1)(1+\phi_2)(1+\delta)}.$$

Now we need to verify that  $x \leq a$ , as conjectured. This can be simplified to obtain

$$a \geq \frac{(1+\phi_1)(1+\phi_2 - \phi_2 b)}{(1+\phi_2)(1+\delta + \phi_1\delta)}.$$

Now we need to verify that  $1-y \leq a$ . This can be simplified to obtain

$$a \geq \frac{\delta(1+\phi_1)(1+\phi_2 - \phi_2 b)}{(1+\phi_2)(1+\delta + \phi_1\delta)}.$$

Note that

$$\frac{\delta(1+\phi_1)(1+\phi_2 - \phi_2 b)}{(1+\phi_2)(1+\delta + \phi_1\delta)} < \frac{(1+\phi_1)(1+\phi_2 - \phi_2 b)}{(1+\phi_2)(1+\delta + \phi_1)}$$

can be simplified to obtain  $\delta^2 < 1$ , which is true. Therefore, the binding condition among these two is that

$$a \geq \frac{(1+\phi_1)(1+\phi_2 - \phi_2 b)}{(1+\phi_2)(1+\delta + \phi_1)}.$$

Next, we need to verify that  $y \leq b$ . This can be simplified to obtain

$$a \geq \frac{(1+\phi_1)(1+\phi_2 - b\delta - b - \phi_2 b)}{\phi_1(1+\phi_2)}.$$

Now we need to verify that  $1-x \leq b$ . This can be simplified to obtain

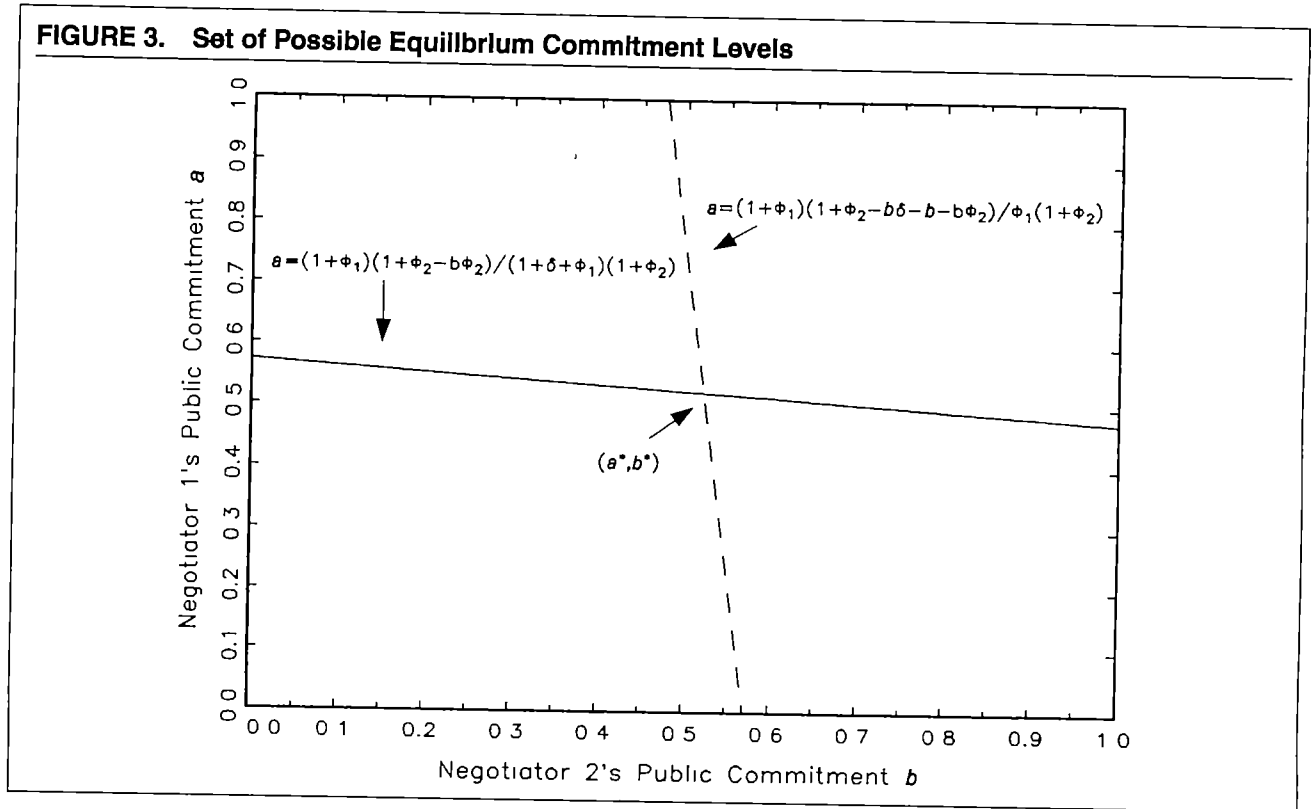
$$a \geq \frac{(1+\phi_1)(\delta + \delta\phi_2 - b\delta - \delta\phi_2 b - b)}{\phi_1\delta(1+\phi_2)}.$$

Note that

$$\begin{aligned} & \frac{(1+\phi_1)(1+\phi_2 - b\delta - b - \phi_2 b)}{\phi_1(1+\phi_2)} \\ & > \frac{(1+\phi_1)(\delta + \delta\phi_2 - b\delta - \delta\phi_2 b - b)}{\phi_1\delta(1+\phi_2)} \end{aligned}$$



**FIGURE 3. Set of Possible Equilibrium Commitment Levels**



can be simplified to obtain  $\delta^2 < 1$ , which is true. Therefore,

$$a \geq \frac{(1 + \phi_1)(1 + \phi_2 - b\delta - b - \phi_2 b)}{\phi_1(1 + \phi_2)}$$

is the binding condition between these two.

Therefore, this SPE of the bargaining subgame exists for all  $(a, b) \in [0, 1] \times [0, 1]$ , such that

$$a \geq \frac{(1 + \phi_1)(1 + \phi_2 - \phi_2 b)}{(1 + \phi_2)(1 + \delta + \phi_1)}$$

and

$$a \geq \frac{(1 + \phi_1)(1 + \phi_2 - b\delta - b - \phi_2 b)}{\phi_1(1 + \phi_2)}.$$

The set of values of  $(a, b)$  such that these two conditions hold consists of the upper right quadrant of Figure 3 (i.e., the region above both of the lines, including the lines themselves).

Note that the two lines in Figure 3 intersect at

$$(a^*, b^*) = \left( \frac{1 + \phi_1}{1 + \delta + \phi_1 + \phi_2}, \frac{1 + \phi_2}{1 + \delta + \phi_1 + \phi_2} \right).$$

For any point in the upper right quadrant of Figure 3, negotiator 1's SPE expected payoff is as follows:

$$\begin{aligned} V_1(a, b) &= p[x - \phi_1(a - x)] + (1 - p)[(1 - y) - \phi_1(a - (1 - y))] \\ &= \frac{[p(1 - \delta) + \delta][1 + \phi_1 + \phi_2 + \phi_1\phi_2 - \phi_2b - \phi_1\phi_2b - \phi_1a - \phi_1\phi_2a]}{(1 + \delta)(1 + \phi_2)} \end{aligned}$$

Note that  $V_1$  is strictly decreasing in  $a$ .

Similarly, negotiator 2's expected payoff is

$$\begin{aligned} V_2(a, b) &= (1 - p)[y - \phi_2(b - y)] + p[(1 - x) - \phi_2(b - (1 - x))] \\ &= \frac{[1 - p(1 - \delta)][1 + \phi_1 + \phi_2 + \phi_1\phi_2 - \phi_1a - \phi_1\phi_2a - \phi_2b - \phi_1\phi_2b]}{(1 + \delta)(1 + \phi_1)}. \end{aligned}$$

Note that  $V_2$  is strictly decreasing in  $b$ .

Thus, since each player's payoff is strictly decreasing in its public commitment, the only possible equilibrium level of public commitments in the upper right quadrant of Figure 3 consists of the lower boundary of this quadrant (i.e., the actual lines). For any other point in the upper right quadrant, each player can strictly increase its payoff by choosing a lower commitment.

We now show that the point of intersection of the two lines

$$(a^*, b^*) = \left( \frac{1 + \phi_1}{1 + \delta + \phi_1 + \phi_2}, \frac{1 + \phi_2}{1 + \delta + \phi_1 + \phi_2} \right)$$

is an equilibrium level of public commitments. (To derive the values of  $x^*$  and  $y^*$  given in Proposition 1, just plug  $a^*$  and  $b^*$  into the formulas for  $x$  and  $y$  derived earlier.) Note that

$$V_1(a^*, b^*) = \frac{[p(1 - \delta) + \delta](1 + \phi_1)}{1 + \delta + \phi_1 + \phi_2}$$

and

$$V_2(a^*, b^*) = \frac{[1 - p(1 - \delta)](1 + \phi_2)}{1 + \delta + \phi_1 + \phi_2}.$$

If a player deviates by choosing a higher public commitment, its payoff decreases. Now we need to verify that a player cannot increase its payoff by choosing a lower public commitment. We only need to show that one player, say player 2, cannot increase its payoff by committing to less than  $b^*$  when player 1 is choosing  $a^*$ . The argument for player 1 is exactly analogous.

Our strategy is to identify a SPE of the bargaining subgame when player 1 is choosing  $a^*$  and player 2 is choosing  $b < b^*$ , for all such possible values of  $b$ . We then show that player 2 is strictly worse off in these equilibria of the bargaining subgame than he is by choosing  $b^*$ . It turns out that there

are two cases that we need to consider: (1) when  $a^* = \frac{1+\phi_1}{1+\delta+\phi_1+\phi_2} < \frac{\delta}{1+\delta}$ , and (2) when  $a^* \geq \frac{\delta}{1+\delta}$ . (Note that  $a^* \leq \frac{\delta}{1+\delta}$  implies that  $b^* > \frac{\delta}{1+\delta}$ , and vice-versa. That is, it is impossible for  $a^* \leq \frac{\delta}{1+\delta}$  and  $b^* \leq \frac{\delta}{1+\delta}$  simultaneously, which would mean that the proposals in the Rubinstein (1982) model would be enough to satisfy each side's public commitment. It is certainly possible for  $a^* > \frac{\delta}{1+\delta}$  and  $b^* > \frac{\delta}{1+\delta}$  simultaneously, for example, when  $\phi_1 = \phi_2$ .)

**Case 1:  $a^* < \frac{\delta}{1+\delta}$ .**

STEP 1(a): Suppose player 2 chooses a slightly lower commitment than  $b^*$ . Then we conjecture that there exists a stationary SPE of the bargaining subgame in which  $x \geq a$ ,  $1-y \leq a$ ,  $y \geq b$ , and  $1-x \leq b$ . The equations for such an equilibrium are

$$\begin{aligned}(1-x) - \phi_2(b - (1-x)) &= \delta y \\ (1-y) - \phi_1(a - (1-y)) &= \delta x.\end{aligned}$$

Solving this pair of simultaneous equations for  $x$  and  $y$ , we obtain

$$\begin{aligned}x &= \frac{1 + \phi_1\phi_2 - b\phi_2 + \delta a\phi_1 - \delta\phi_1 + \phi_1 + \phi_2 - \delta - \phi_1\phi_2b}{1 + \phi_1 + \phi_2 + \phi_1\phi_2 - \delta^2} \\ y &= \frac{1 + \phi_1\phi_2 - a\phi_1 + \delta b\phi_2 - \delta\phi_2 + \phi_1 + \phi_2 - \delta - \phi_1\phi_2a}{1 + \phi_1 + \phi_2 + \phi_1\phi_2 - \delta^2}.\end{aligned}$$

Now we need to verify that  $x \geq a$ . Substituting  $a^*$  for  $a$ , this can be simplified to obtain  $b \leq \frac{1+\phi_2}{1+\delta+\phi_1+\phi_2}$ .  $y \geq b$  can be simplified to obtain the same thing.  $1-x \leq b$  can be simplified to obtain  $b \geq \frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}$ .  $1-y \leq a$  can be simplified to obtain

$$b \geq \frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta}.$$

Setting

$$\begin{aligned}&\frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta} \\ &> \frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}\end{aligned}$$

and simplifying, we obtain  $a^* < \frac{\delta}{1+\delta}$ , which we have supposed to be true in this case. Setting

$$\begin{aligned}&\frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta} \\ &< \frac{1+\phi_2}{1+\delta+\phi_1+\phi_2}\end{aligned}$$

and simplifying, we obtain  $\delta < 1$ , which is true. Therefore, the binding condition for this SPE of the bargaining subgame to exist is

$$\begin{aligned}&\frac{1+\phi_2}{1+\delta+\phi_1+\phi_2} \\ &\geq b \geq \frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta}.\end{aligned}$$

In this SPE, negotiator 2's expected payoff is  $V_2(a, b) = (1-p)y + p[(1-x) - \phi_2(b - (1-x))] = [1-p(1-$

$\delta)]y$ . Looking at  $y$ , we see that  $V_2(a, b)$  is strictly increasing in  $b$ . And at the upper bound of this equilibrium,  $V_2(a^*, b^*) = \frac{[1-p(1-\delta)](1+\phi_2)}{1+\delta+\phi_1+\phi_2}$ . Therefore, player 2 cannot profitably deviate to this SPE of the bargaining subgame.

STEP 1(b): Now suppose player 2 chooses an even lower commitment. Then we conjecture that there exists a stationary SPE of the bargaining subgame in which  $x$ ,  $1-y \geq a$  and  $y \geq b$  and  $1-x \leq b$ . The equations for such an equilibrium are

$$\begin{aligned}(1-x) - \phi_2(b - (1-x)) &= \delta y \\ 1-y &= \delta x.\end{aligned}$$

Solving this pair of simultaneous equations for  $x$  and  $y$ , we obtain

$$\begin{aligned}x &= \frac{1 - \phi_2b + \phi_2 - \delta}{1 + \phi_2 - \delta^2} \\ y &= \frac{1 + \phi_2 - \delta + \phi_2b\delta - \delta\phi_2}{1 + \phi_2 - \delta^2}.\end{aligned}$$

Now we need to verify that  $x \geq a$ . Substituting  $a^*$  for  $a$  and simplifying, we obtain

$$b \leq \frac{\phi_2 + \phi_2^2 - \delta\phi_1 + \delta^2\phi_1}{\phi_2(1+\delta+\phi_1+\phi_2)}.$$

$y \geq b$  can be simplified to obtain

$$b \leq \frac{1 + \phi_2 - \delta - \delta\phi_2}{1 + \phi_2 - \delta^2 - \delta\phi_2}.$$

$1-y \geq a$  can be simplified to obtain

$$b \leq \frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta}.$$

Note that

$$\begin{aligned}&\frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta} \\ &< \frac{1 + \phi_2 - \delta - \delta\phi_2}{1 + \phi_2 - \delta^2 - \delta\phi_2}\end{aligned}$$

as well as

$$\begin{aligned}&\frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta} \\ &< \frac{\phi_2 + \phi_2^2 - \delta\phi_1 + \delta^2\phi_1}{\phi_2(1+\delta+\phi_1+\phi_2)}\end{aligned}$$

can be simplified to obtain  $\delta < 1$ , which is true. Therefore,

$$b \leq \frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta}$$

is the binding condition among these three.

Finally,  $1-x \leq b$  can be simplified to obtain  $b \geq \frac{\delta}{1+\delta}$ . Note that

$$\begin{aligned}&\frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta} \\ &> \frac{\delta}{1+\delta}\end{aligned}$$

can be simplified to obtain  $a^* < \frac{\delta}{1+\delta}$ , which we have supposed to be true for this case. Therefore, this SPE of the bargaining subgame exists for all

$$\frac{\phi_1\phi_2\delta + \delta + 2\phi_2\delta + \phi_2^2\delta + \phi_1\delta - \delta^3 - 1 - \phi_1 - \phi_2 + \delta^2 - \phi_1\phi_2}{\phi_2^2\delta + \phi_2\delta + \phi_2\delta^2 + \phi_1\phi_2\delta} \geq b \geq \frac{\delta}{1+\delta}.$$

In this SPE, negotiator 2's expected payoff is  $V_2(a, b) = (1-p)y + p[(1-x) - \phi_2(b - (1-x))] = [1 - p(1-\delta)]y$ . Looking at  $y$ , we see that  $V_2(a, b)$  is strictly increasing in  $b$ . And at the upper bound of this equilibrium, we are at the lower bound of the previous equilibrium, which we already know is strictly worse for player 2 than is  $V_2(a^*, b^*)$ . Therefore, player 2 cannot profitably deviate to this SPE of the bargaining subgame.

STEP 1(c): Now suppose player 2 chooses some  $b \leq \frac{\delta}{1+\delta}$ . Then each side's proposal in the Rubinstein (1982) model is enough to satisfy each side's commitment; hence, the SPE of the Rubinstein model is the SPE of this case. Negotiator 2's payoff in this range does not depend on  $b$ , and his payoff is equal to his payoff at the lower bound of the previous equilibrium, which we already know is strictly worse for player 2 than is  $V_2(a^*, b^*)$ . Therefore, player 2 cannot profitably deviate to this SPE of the bargaining subgame.

STEP 1(d): Therefore, we have shown that when  $a^* < \frac{\delta}{1+\delta}$ , player 2 is strictly worse off by choosing any  $b < b^*$ .

**Case 2:  $a^* \geq \frac{\delta}{1+\delta}$ .**

STEP 2(a): Suppose player 2 chooses a slightly lower commitment than  $b^*$ . Then the same SPE in step 1(a) exists (the argument is exactly the same as there), only now the binding condition is

$$\frac{1+\phi_2}{1+\delta+\phi_1+\phi_2} \geq b \geq \frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}.$$

(Note that  $\frac{1+\phi_2}{1+\delta+\phi_1+\phi_2} > \frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}$  can be simplified to obtain  $\phi_1+\phi_2+\phi_1\phi_2 > \delta^2-1$ , which is true, and therefore this range for  $b$  exists.) In step 1(a), we showed that player 2 cannot profitably deviate to this SPE of the bargaining subgame.

STEP 2(b): Suppose player 2 chooses an even lower commitment. Then we conjecture that there exists a stationary SPE of the bargaining subgame in which  $x \geq a$  and  $1-y \leq a$  and  $y, 1-x \geq b$ . The equations for such an equilibrium are

$$\begin{aligned} 1-x &= \delta y \\ (1-y) - \phi_1(a - (1-y)) &= \delta x. \end{aligned}$$

Solving this pair of simultaneous equations for  $x$  and  $y$ , we obtain

$$\begin{aligned} y &= \frac{(1-\delta) + \phi_1(1-a)}{(1+\phi_1-\delta^2)} \\ x &= \frac{(1+\phi_1)(1-\delta) + \phi_1\delta a}{(1+\phi_1-\delta^2)}. \end{aligned}$$

Now we need to verify that  $x \geq a$ . Substituting  $a^*$  for  $a$ , this can be simplified to obtain  $\phi_2 \geq 0$ , which is true.  $1-y \leq a$  can be simplified to obtain  $a^* \geq \frac{\delta}{1+\delta}$ , which we have supposed to be true in this case.  $1-x \geq b$  can be simplified to obtain

$$b \leq \frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}.$$

Finally,  $y \geq b$  can be simplified to obtain

$$b \leq \frac{1+\phi_1+\phi_2-\delta^2-\delta\phi_2+\phi_1\phi_2}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}.$$

Note that

$$\begin{aligned} &\frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)} \\ &< \frac{1+\phi_1+\phi_2-\delta^2-\delta\phi_2+\phi_1\phi_2}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)} \end{aligned}$$

can be simplified to obtain  $\delta < 1$ , which is true. Therefore, this SPE exists for all

$$b \leq \frac{\delta(\phi_1+1+\phi_1\phi_2-\phi_2\delta-\delta^2+\phi_2)}{(1+\delta+\phi_1+\phi_2)(1+\phi_1-\delta^2)}.$$

In this SPE, negotiator 2's expected payoff is  $V_2(a, b) = (1-p)y + p(1-x) = [1 - p(1-\delta)]y$ . Looking at  $y$ , we see that  $V_2(a, b)$  does not depend on  $b$ , and negotiator 2's payoff is the same as at the lower bound of the previous equilibrium, which we already know is strictly worse for player 2 than is  $V_2(a^*, b^*)$ . Therefore, player 2 cannot profitably deviate to this SPE of the bargaining subgame.

STEP 2(c): Therefore, we have shown that when  $a^* \geq \frac{\delta}{1+\delta}$ , player 2 is strictly worse off by choosing any  $b < b^*$ . ■

## REFERENCES

- Binmore, Ken G. 1987. "Perfect Equilibria in Bargaining Models." In *The Economics of Bargaining*, ed. Ken G. Binmore and Partha Dasgupta. Oxford: Basil Blackwell.
- Crawford, Vincent P. 1982. "A Theory of Disagreement in Bargaining." *Econometrica* 50 (May): 607-38.
- de Tocqueville, Alexis. [1835] 1945. *Democracy in America*. New York: Knopf.
- Fearon, James. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (September): 577-92.
- Fearon, James. 1995. "Rationalist Explanations for War." *International Organization* 49 (Summer): 379-414.
- Fearon, James. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41 (February): 68-90.
- Gowa, Joanne. 1995. "Democratic States and International Disputes." *International Organization* 49 (Summer): 511-22.
- Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (April): 175-200.
- Iida, Keisuke. 1993. "When and How Do Domestic Constraints Matter? Two-Level Games with Uncertainty." *Journal of Conflict Resolution* 37 (September): 403-26.
- Jordan, Amos A., William J. Taylor, and Michael J. Mazarr. 1999. *American National Security*. 5th ed. Baltimore: Johns Hopkins University Press.
- Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.
- Kydd, Andrew, and Barbara Walter. 2002. "Sabotaging the Peace: The Politics of Extremist Violence." *International Organization* 56 (Spring): 263-96.
- Makovsky, David. 2001. "Middle East Peace Through Partition." *Foreign Affairs* (March/April): 28-45.
- Martin, Lisa L. 2000. *Democratic Commitments: Legislatures and International Cooperation*. Princeton: Princeton University Press.
- Milner, Helen V. 1997. *Interests, Institutions, and Information: Domestic Politics and International Relations*. Princeton: Princeton University Press.
- Mo, Jongryn. 1994. "The Logic of Two-Level Games with Endogenous Domestic Coalitions." *Journal of Conflict Resolution* 38 (September): 402-22.

- Mo, Jongryn. 1995. "Domestic Institutions and International Bargaining: The Role of Agent Veto in Two-level Games." *American Political Science Review* 89 (December): 914–24.
- Morgenthau, Hans. 1956. *Politics Among Nations*. 2d ed. New York: Knopf.
- Muthoo, Abhinay. 1992. "Revocable Commitment and Sequential Bargaining." *Economic Journal* 102 (March): 378–87.
- Muthoo, Abhinay. 1996. "A Bargaining Model Based on the Commitment Tactic." *Journal of Economic Theory* 69 (April): 134–52.
- Muthoo, Abhinay. 1999. *Bargaining Theory with Applications*. Cambridge: Cambridge University Press.
- Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18 (April): 155–62.
- Pahre, Robert. 1997. "Endogenous Domestic Institutions in Two-Level Games and Parliamentary Oversight of the European Union." *Journal of Conflict Resolution* 41 (February): 147–74.
- Perlmutter, Amos. 1995. "The Israeli-PLO Accord Is Dead." *Foreign Affairs* (May/June): 59–68.
- Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42 (Summer): 427–60.
- Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50 (January): 79–109.
- Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (Winter): 121–49.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
- Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform?" *International Organization* 53 (Spring): 233–66.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (September): 623–38.
- Telhami, Shibley. 1990. *Power and Leadership in International Bargaining: The Path to the Camp David Accords*. New York: Columbia University Press.
- Waltz, Kenneth. 1979. *Theory of International Politics*. New York: McGraw-Hill.

# Nested Analysis as a Mixed-Method Strategy for Comparative Research

EVAN S. LIEBERMAN *Princeton University*

**D**espite repeated calls for the use of “mixed methods” in comparative analysis, political scientists have few systematic guides for carrying out such work. This paper details a unified approach which joins intensive case-study analysis with statistical analysis. Not only are the advantages of each approach combined, but also there is a synergistic value to the nested research design: for example, statistical analyses can guide case selection for in-depth research, provide direction for more focused case studies and comparisons, and be used to provide additional tests of hypotheses generated from small-N research. Small-N analyses can be used to assess the plausibility of observed statistical relationships between variables, to generate theoretical insights from outlier and other cases, and to develop better measurement strategies. This integrated strategy improves the prospects of making valid causal inferences in cross-national and other forms of comparative research by drawing on the distinct strengths of two important approaches.

**L**ong-standing methodological debates highlighting inherent tradeoffs in the main modes of comparative analysis have tended to force scholars to choose between one of two imperfect approaches. On the one hand, even while defending its merits, Lijphart (1971, 685) succinctly identified the central shortcoming of the “comparative method” as the problem of “many variables, small number of cases.” In the years to follow, some scholars argued that such attempts to draw general conclusions from intensive analysis of one or a few cases have been flawed by various problems of selection bias, lack of systematic procedures, and inattention to rival explanations (e.g., Achen and Snidal 1989; Geddes 1990; King, Keohane, and Verba 1994). Alternatively, other scholars have argued not only that some of the critiques of qualitative research may be overdrawn and the contributions of these works underappreciated, but also that the complex phenomena and causal processes associated with big, national-level outcomes require a more close-range analytic tool that is less likely to generate spurious results (e.g., Collier, Brady, and Seawright 2004; Collier and Mahoney 1996; Munck 1998; Rogowski 1995). Qualitatively oriented scholars have their own tradition of challenging the statistical approach, including Sartori’s (1970) powerful invective against “conceptual stretching,” which in turn has been refuted by scholars such as Jackman (1985), who argues that the comparative method is a “weak approximation of the statistical method,” (165) and that “cross-national statistical analyses have a lot to offer” (179).

Although such back-and-forth debate has served to illuminate the shortcomings in various methodological approaches, it has also provided momentum for greater synthesis of research styles and findings. Two decades after publication of Lijphart’s (1971) article, Collier (1991) pointed out that advances in both statistical and small-N approaches, and evidence of increasing communication across the two approaches, held great promise for scholarly progress. Both King, Keohane, and Verba’s *Designing Social Inquiry* (1994) and Brady and Collier’s *Rethinking Social Inquiry* (2004) have demonstrated that each mode of analysis can be successfully used to achieve similar social scientific ends, while using somewhat different tools. And yet, these contributions have largely assumed that there will continue to be substantial divisions of scholarly labor, even as research findings across the methodological divide are often ignored. In a somewhat different formulation, several scholars have called for greater *integration* of methodological approaches (Achen and Snidal 1989; Tarrow 1995) or the mixing of methods. Despite the initially appealing nature of such a resolution, scholars have received little guidance about how to blend these modes of analysis. As Bennett (2002) points out in a paper reviewing some of the ways in which case study, statistical, and formal methods have been combined in political science, there is a need to focus on the ways in which such combinations could be increased and improved. Clearly, not all forms of mixed strategies will provide greater insights into particular research problems. In fact, some may simply generate more confusion than clarity.

This article systematizes a unified “mixed method” approach to comparative research, which I call *nested analysis*.<sup>1</sup> It combines the statistical analysis of a large

Evan S. Lieberman is Assistant Professor of Politics at Princeton University, 239 Corwin Hall, Princeton, NJ 08544 (ESL@Princeton.edu).

Earlier versions of this paper were presented at the 2002 Annual Meeting of the American Political Science Association, Boston, MA, and at a 2004 Workshop on multilevel models at Princeton University. In addition, Larry Bartels, Andrew Bennett, Nancy Bermeo, David Collier, Michael Coppedge, John Gerring, Marc Morjé Howard, David Laitin, Phil Shively, students in my Qualitative Methods graduate seminars at Princeton, members of the Princeton Comparative Politics Working Group, three anonymous reviewers and the editor of the *APSR* all provided extremely helpful comments and suggestions.

<sup>1</sup> In this article, I discuss Coppedge’s (2005) use of what he calls “nested inference” in an analysis of the breakdown of democracy in Venezuela. Although he is methodologically self-conscious in describing how case study and quantitative/large sample analyses are combined in an application, that study represents one variation of the approach I describe in this article, which attempts to anticipate and systematize a broader range of research problems and strategies.

sample of cases with the in-depth investigation of one or more of the cases contained within the large sample. This would include the study of a nation-state nested within an analysis of 50 nation-states; the study of two provinces nested within an analysis of 20 provinces; or the study of an institution nested within an analysis of 100 institutions. Although all of the examples discussed in the article are concerned with country- or national-level analyses, the strategies described here should apply to any comparative analysis of social units for which both quantitative and in-depth case study data can be obtained. Thus, the approach could be applied to the analysis of individual behaviors or attitudes, but only if the researcher were willing and able to gather new data about particular individuals through intensive interview or related approaches in combination with quantitative analyses of large-scale surveys. If the study concerned specific, well-studied individuals, such as presidents or legislators, for which additional information could be gleaned from in-depth research of particular cases, the approach described here would indeed apply.<sup>2</sup>

I should be clear that the strategy described here is quite distinct from the message outlined by King, Keohane, and Verba (1994). Rather than advocating that there are “lessons” useful for qualitative researchers that can be gleaned from the logic of statistical analysis (or *vice-versa*, an argument they do not make) I show that there are specific benefits to be gained by deploying *both* analytical tools simultaneously, and I emphasize the benefits of distinct complementarities rather than advocating a single style of research. Although the move from “small-N” analysis (SNA) to nested analysis obviously requires that one “find additional cases” (King, Keohane, and Verba 1994, 208–29), it assumes that it may be extremely difficult and inefficient to gather perfectly equivalent data for each case, and that the inferential opportunities from the “large-N” analysis (LNA) will be distinctive.<sup>3</sup>

## OVERVIEW OF THE NESTED ANALYSIS APPROACH

I describe a set of strategies for gaining maximum analytic leverage when combining SNA and LNA within a single framework (summarized in Figure 1). Although there is an enormous variety of analytical strategies contained under these two subheadings, both in terms of actual number of units analyzed and the scope of the time dimension considered, for the purposes of this

paper, it is useful to make a general distinction: I define LNA as a mode of analysis in which the primary causal inferences are derived from statistical analyses which ultimately lead to quantitative estimates of the robustness of a theoretical model; I define SNA as a mode of analysis in which causal inferences about the primary unit under investigation are derived from qualitative comparisons of cases and/or process tracing of causal chains within cases across time, and in which the relationship between theory and facts is captured largely in narrative form.<sup>4</sup> The strategy of combining the two approaches aims to improve the quality of conceptualization and measurement, analysis of rival explanations, and overall confidence in the central findings of a study. The promise of the nested research design is that both LNA and SNA can inform each other to the extent that the analytic payoff is greater than the sum of the parts. Not only is the information gleaned complementary, but also each step of the analysis provides direction for approaching the next step. Most prominently, LNA provides insights about rival explanations and helps to motivate case selection strategies for SNA, whereas SNA helps to improve the quality of measurement instruments and model specifications used in the LNA.

As a thumbnail sketch, the approach involves starting with a preliminary LNA and making an assessment of the robustness of those results. If the model is well specified and the results are robust, one proceeds to “Model-testing Small-N Analysis,” and if not, to “Model-building Small-N Analysis.” In each case, as shown in Figure 1, the analyst must again make assessments about the findings from such analysis, using directions and insights gleaned from the SNA, and those assessments provide a framework for either ending the analysis or carrying out additional iterations of SNA or LNA. Detailing the nature of the particular strategies for carrying out each type of analysis, as well as the nature of the assessments, is the central goal of the remainder of the paper.

Nested analysis is resolutely “catholic” in its assumptions and objectives. It assumes an interest in *both* the exploration of general relationships and explanations and the specific explanations of individual cases and groups of cases. For example, a nested research design implies that scholars will pose questions in forms such as “What causes social revolutions?” while simultaneously asking questions such as “What was the cause of social revolution in France?” Nested analysis helps scholars to ask good questions when analyzing their data and to be resourceful in finding answers.

Before proceeding to detail the procedures associated with nested analysis, it is important to situate the strategy within the context of two other proposals for “alternative” methodological approaches. First, Charles Ragin (1987, 2000) attempts to steer a middle path between “quantitative” and “qualitative”

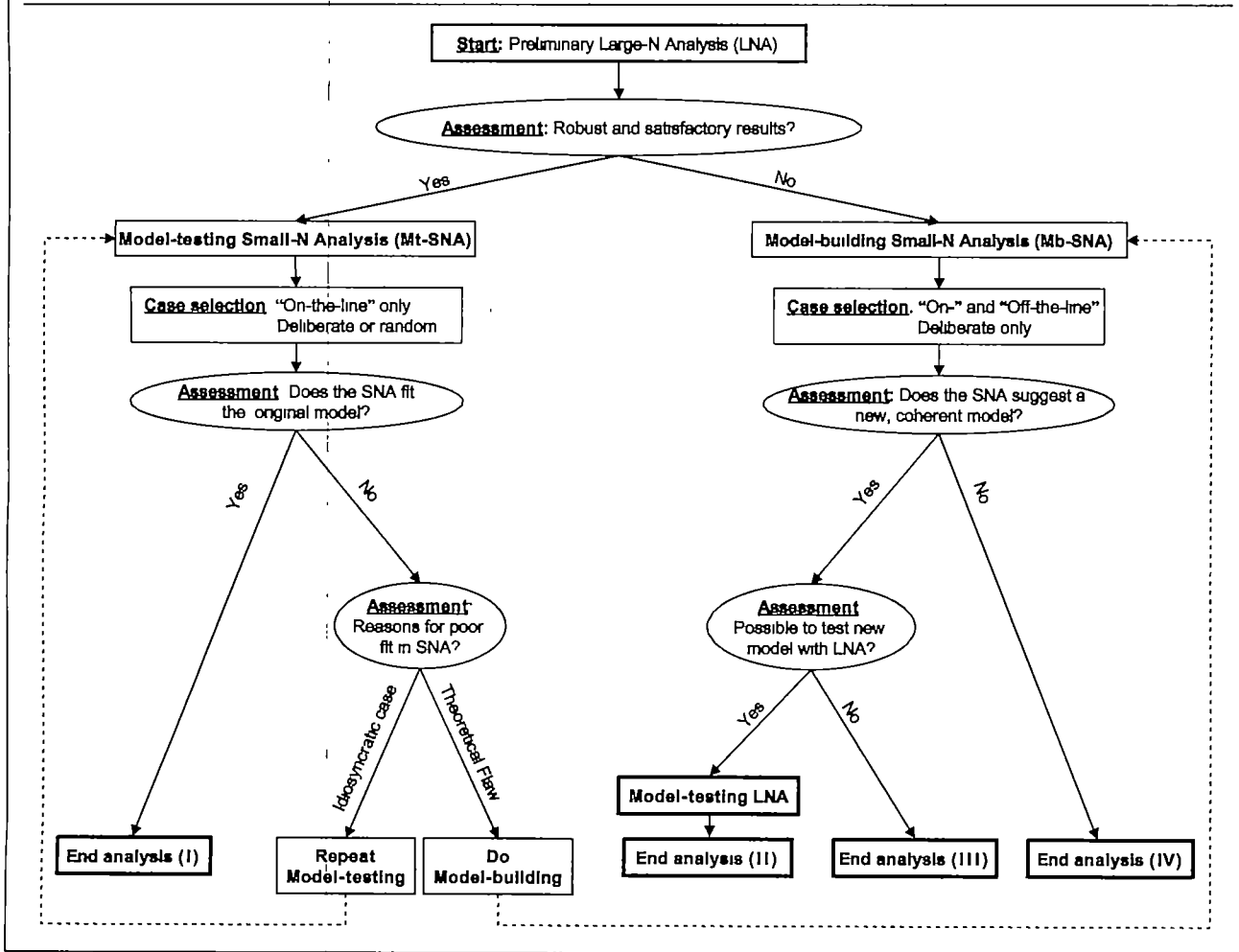
<sup>2</sup> However, for most analyses of individual behaviors or attitudes, for which the “large-N” component of the data is contained in a survey, I would not expect this approach to be feasible, because scholars are unlikely to be able to conduct further in-depth research with the original respondents. Moreover, the prospect of explaining the exceptional nature of a particular individual is unlikely to be of intrinsic interest in the way scholars are likely to be interested in the particularities of larger social units, such as national states.

<sup>3</sup> By “cases” I mean the shared unit of analysis. In cross-national research, each case is a country.

<sup>4</sup> As is discussed in the text, qualitative analysis is the hallmark, but not the defining feature of SNA. Within-case analyses may include a range of statistical analyses of data that are not available across the larger sample of primary unit cases (i.e., countries).



FIGURE 1. Overview of the Nested Analysis Approach



research with his specification of a Boolean approach and elaboration of a "fuzzy set"/Qualitative Comparative Analysis (FsQCA). Ultimately, his strategy focuses on integrating close-range analysis to ensure the proper delineation of theoretically relevant populations and valid classification of cases, with an algorithm that finds the necessary and sufficient conditions associated with particular sets of phenomena. Second, the Bayesian approach (Western and Jackman 1994), like the FsQCA approach, and distinct from the classical regression model, relies heavily on investigator knowledge of cases and processes, but does this through the formal introduction of subjective probability estimates. However, neither the stated approaches to Bayesian analysis nor FsQCA provide direction about how to gather additional research in the SNA—they assume a seamless discovery process of "outside knowledge," with almost no focus on the specific role of gathering and reporting case materials. In making prescient critiques of standard cross-country regression analyses, advocates of both the Bayesian and the FsQCA approaches allow for the inductive incorporation of knowledge from cases, but as currently formulated, they provide little guidance about the cases we should study or what role they ought to play in the assess-

ment of theoretical findings.<sup>5</sup> As such, both of these approaches may serve as partial correctives to cross-country regression analysis, but neither is complete. For the purposes of nested analysis, both FsQCA and Bayesian approaches may be used in the LNA, and the guidelines developed here for combining such approaches with SNA should still apply.

It is also important to indicate that the nested analysis approach is agnostic with respect to the source of theory formation. Although others have explicitly included the development of formal—that is, mathematically specified—theory in their discussions and proposals for integrating approaches to the study of comparative politics (Bates et al. 1998; Laitin 2002), the nested analysis approach has no particular affinity for any single theoretical approach, except for a more general positivist goal of causal inference. Such theory may be developed and conveyed in a nonmathematical form (i.e., "No bourgeoisie, no democracy") or through the use of mathematical operators and proofs. Along these lines, the nested analysis approach allows for both

<sup>5</sup> Certainly, the nested analysis approach could be described as a "folk Bayesian" approach (McKeown 2004, 158–62) in that it seeks to formally introduce investigator knowledge of the world.

the testing of deductively formed hypotheses and the inductive generation of theory. Many of the benefits of nested analysis explicitly rest on the assessment that the overall state of theory in cross-national research is relatively thin with respect to the questions being asked, and that empirical analysis is required both to develop hypotheses and to test them. Of course, the nature of the specific hypotheses—including the reliance on microfoundational or macrostructural mechanisms—is likely to shape the evidentiary requirements, particularly in the SNA (as discussed later). In the remainder of the article, my use of the term *model* implies only a general theoretical argument that relates explanatory to outcome variables, and should not be taken to imply a “formal” model.

The central objective of the remainder of the article is to specify a set of procedures for integrating LNAs and SNAs. Although it is neither possible nor desirable to identify a cookie-cutter approach to analysis, the systematization of these steps should provide a clear logic for integrating the two types of analyses and for identifying the types of assumptions and justifications that are required for analysis. As always, scholarly tastes and subjective judgments about the robustness of the results influence how the nested analysis will proceed,<sup>6</sup> but it is important to ensure a high degree of transparency, particularly when adding complexity to the scope of analysis.

I use examples of published and unpublished studies to demonstrate the use of various techniques within the larger approach of nested analysis, but the article is not intended to be a review of the literature as much as an outline for the execution of nested analysis. Indeed, because almost none of the examples actually employ the specific language or framework developed here, I only claim that these examples help to clarify how aspects of the approach have been used in particular studies and with what benefit.

## STARTING THE ANALYSIS: PRELIMINARY LNA

Scholars engage new research projects with varying levels of background information about a specific case or set of cases, but the nested analysis formally *begins* with a quantitative analysis, or preliminary LNA. Thus, a prerequisite for carrying out a nested analysis is availability of a quantitative dataset, with a sufficient number of observations for statistical analysis,<sup>7</sup> and a baseline theory. The preliminary LNA provides information that should ultimately complement the findings of the SNA, and that will guide the execution of the SNA. Particularly for scholars who would

have carried out SNA exclusively, the preliminary LNA requires explicit consideration of the universe of cases for which the theory ought to apply, and identification of the range of variation on the dependent variable. It also provides opportunities to generate clear baseline estimates of the strength of the relationship between variables, including estimates of how confident we can be about those relationships given a set of assumptions about probabilities and frequencies. When scholars begin with strong hypotheses and good data, the preliminary LNA can be understood as a more conventional hypothesis-testing analysis.

The content of the LNA may take one of several forms, depending on the availability of data, and the nature of the causal model—for example, depending on whether the outcome is understood to be graded or dichotomous, and whether the hypothesized relationship is understood in probabilistic or deterministic terms. One may use multivariate regression analysis; fuzzy set/qualitative comparative analysis (FsQCA); bivariate/correlational analysis, or simply descriptive statistics to analyze the scores on the dependent variable. Decisions about which brand of analysis to use, and the nature of the model—linear or curvilinear, for example—must be made with respect to available data and theory. In any case, the goal of the preliminary LNA is to explore as many appropriate, testable hypotheses as is possible with available theory and data. Indeed, the very feasibility of nested analysis is a product of the increasing availability of datasets produced by other scholars and international organizations, obviating the need for independent data collection, at least at this preliminary stage. (Significant independent collection of data at this stage can be justified only when a scholar has very strong initial hypotheses and great confidence in how to measure key variables.) However, it is important to note that the preliminary LNA should avoid the insertion of any control variables that do not have a clear theoretical justification, such as regional “dummy” variables. Such variables *are* likely to soak up some of the cross-country variance, leaving less to be explained in the SNA, but in the absence of good theory, such controls weigh against the nested approach, which aims to answer the very question of why groups of countries might vary in systematic ways.

A core strength of LNA relative to SNA is its ability to simultaneously estimate the effects of rival explanations and/or control variables on an outcome of interest. To a large extent, SNA in the field of comparative politics has relied on variants of Mill’s methods in order to deal with country-level rival explanations—that is, scholars identify cases that score similarly on several key variables, using shared traits as a basis for analytical equivalence approximating statistical control.<sup>8</sup> Although the strategy of identifying cases with relatively similar scores on such variables can be a powerful one, in a nonexperimental setting, important differences among cases can almost always be identified,

<sup>6</sup> Indeed, there is no consensus about the robustness of a particular  $R^2$  statistic, or what amount of process-tracing evidence should be considered persuasive

<sup>7</sup> There is no clear lower bound for the number of cases that can be analyzed through a statistical analysis, but fewer cases obviously reduce the degrees of freedom and intrinsic power of the analysis. It is rare to see quantitative analyses of fewer than 12 cases in cross-country regression analyses

<sup>8</sup> See Gerring 2001 (209–14) for a summary of these methods, often understood as “most similar” and “most different” systems research designs

and these emerge as possible rival explanations. Regardless of whether one's causal model is probabilistic or deterministic, some degree of covariance between a rival explanatory variable and the outcome requires attention within a SNA based on the juxtaposition of "similar" cases. One may attempt to draw on theory to argue why a particular variable is an implausible influence, but skeptics are likely to demand empirical proof. Moreover, one may attempt to carry out "within-case" analysis (Collier 1999) to address rival hypotheses, but again, there may be no over-time variation or other relevant data to analyze; or, one may try to find an additional "similar" case with less variation on the offending variable, but in a world with a limited number of highly heterogeneous countries, such options may be limited. Alternatively, SNA scholars can ignore the cross-case variance or simply concede that there is no way to address the problem with available data. Obviously, these are not ideal solutions.

Depending on the question or the cases under investigation, LNA may be able to lend a hand. Assuming that the LNA is conducted as a regression, the relevant dependent variable can be regressed on measures of the rival explanatory variable under investigation in order to assess the strength of a relationship, particularly when the SNA provides no solid basis for analysis. For example, in her study of multilateral sanctions, Lisa Martin (1992) precedes her analysis of four major case studies with a set of regression analyses, which allows her to assess the general plausibility and implausibility of several possible explanations of why states cooperate to impose economic sanctions. She argues that this technique "has allowed us to narrow the range of hypotheses deserving more-detailed analysis by suggesting that some hypotheses . . . have little empirical support" (92). For example, she finds no support for Keohane's (1984) "declining hegemony" thesis in the LNA (91), which allows her to focus her attention on other possible explanations in the subsequent SNA. In the absence of such LNA, Martin would have been forced to consider Keohane's (1984) important hypothesis in the SNA, imposing analytic costs, and leaving readers to wonder about the weight of this explanation in the larger sample. Alternatively, if the LNA had provided initial support for the Keohane thesis, Martin either would have been forced to accept the usefulness of that model—and perhaps demonstrate that other complementary explanations were possible—or would have been forced to demonstrate in quite convincing terms within the SNA why statistical relationships were likely spurious. Clearly the most powerful refutation of a rival explanation is the presentation of disconfirming evidence in *both* LNA and SNA, but given data and analytic constraints, the ability to rule out a hypothesis in the LNA provides sound justification for focusing on other explanations in the SNA.

At least as important as its ability to dismiss rival explanations, LNA provides a unique instrument for assessing the strength of partial explanations or control variables. Because country-level outcomes tend to be the product of several factors, preliminary LNA is likely to find that some variables *are* significant predic-

tors of the outcome under investigation, even if they can account for only a limited portion of the cross-country variance. For example, in a study of the development of the tax state, Lieberman (2003) begins his analysis by demonstrating that approximately 40% of the cross-national variation in levels of income tax collections is predicted by levels of GDP/capita. Although this variable is essentially treated as a control variable throughout the book, it is extremely useful to have an estimate of the extent to which such a variable helps to explain patterns of variation on the outcome. Much small-N research involves the comparison of "similar" cases. However, because we only observe cases in which there is little to no variation on key control variables, we have little basis for making inferences about the need to control on those variables, or about how strong an influence we should expect those variables to have on the outcomes under investigation. The "puzzle" of a particular case or set of cases can be made clear when we have some estimate of predicted outcomes given a set of parameter estimates and the case scores on those variables.<sup>9</sup>

### ASSESSING THE FINDINGS OF THE LNA: ARE THE RESULTS ROBUST AND SATISFACTORY?

Beyond providing insights into the range of variation on the dependent variable, and estimates of the strength of rival hypotheses and control variables, LNA also provides important information about how to carry out the next stage of the analysis—intensive examination of one or more cases. First, the scholar must assess the findings: did the preliminary LNA provide strong grounds for believing that the initial theoretical model explained the phenomenon being studied?

As noted previously, it is not possible to provide absolute criteria for answering the question about the robustness of the LNA results because subjective assessments about the state of knowledge and what constitutes strong evidence weigh heavily.<sup>10</sup> Depending on the nature of the LNA, standard assessments about the strength of parameter estimates must be used to evaluate goodness of fit between the specified model and the empirical data. Nonetheless, one important tool is central to the nested analysis approach: the actual scores of the cases should be plotted graphically relative to the predicted scores from the statistical estimate,<sup>11</sup> and with proper names attached. This provides an opportunity to make specific assessments of the goodness

<sup>9</sup> Although it is true that these initial parameter estimates are likely to be biased because of model misspecifications, including missing variable bias, our presumption is that when we do not have a fully specified or complete theoretical model, it is useful to gain a sense of what can be explained by the theory and data that *are* available.

<sup>10</sup> For a classic statement on the use of common sense and professional judgment in the use of quantitative analysis, see Achen (1982), especially pp. 29–30.

<sup>11</sup> At the extreme, if no statistical relationship is found between any of the explanatory variables and the outcome of interest, one could simply use a central tendency of the data, such as the mean, as a baseline model, and country cases could be plotted as deviations from the mean.

of model fit with the available cases. In combination with the parameter estimates generated from the LNA, the scholar must decide if the unexplained variance is largely the product of random noise, or if there is reason to believe that a better model/explanation could be formulated. As in any statistical analysis, diagnostic plots may highlight suspect patterns of nonrandom variation in one or more cases—the identification of outliers. However, unlike in surveys of individuals, where case identities are anonymous and thus irrelevant for analysis, in the study of nation-states and many other organization forms, the location of specific cases with respect to the regression line may strongly influence one's satisfaction with the model. For example, a scholar may feel unsatisfied with a model that cannot explain a case perceived to be of great significance within the scholarly literature (e.g., the French revolution in the study of revolutions), or the identification of an outlier case may immediately suggest a new theoretical specification with potentially broader application. If a scholar enters the research project with specific hunches about seemingly anomalous outcomes, analysis of the actual-versus-predicted-scores plot may demonstrate that one or more cases are indeed outliers that may warrant more theoretical attention. Indeed, Lieberman's (2003) study was motivated by a hunch that differences in the Brazilian and South African tax structures were striking and not readily explainable, and the preliminary LNA confirmed that this was true even when key control variables were taken into account. Of course, such preliminary analysis could have served to foreclose unnecessary research by demonstrating that a particular case was (surprisingly) well explained by the existing state of theory.

Using such analyses, the scholar must answer the question: "Were all of the most important hypotheses tested and were the results robust/satisfactory?" The answer to this question informs the approach to the nested case analyses, or SNA, as described in the following section.

### NESTING INTENSIVE CASE STUDIES (SNA) INTO THE ANALYSIS

The second major step of the nested analysis involves the intensive analysis of one or more country cases.<sup>12</sup> Of course, there is nothing particularly distinctive about the simple combination of LNA and SNA; scholars have long recognized the value of "triangulation" for descriptive and causal inference.<sup>13</sup> My contention is that there are several important strategies that can be gleaned from assessment of the LNA, which will narrow the larger menu of options for executing the SNA.<sup>14</sup> Moreover, I emphasize that the best use of

SNA is to leverage its distinct complementarities with LNA, not to try to implement it with the exact same procedures as one would carry out regression analysis. Although many small-N scholars may have an "implicit" regression model in their head when they carry out their analyses, there are clear benefits to being explicit.<sup>15</sup>

It is important to recall that the goal of a nested analysis is ultimately to make inferences about the unit of analysis that is shared between the two types of analysis—typically countries or country-periods. In pursuing this goal, a nested analysis requires a *shifting* of levels of analysis because the SNA component demands an examination of *within*-case processes and/or variation.<sup>16</sup> The SNA should be used to answer those questions left open by the LNA—either because there were insufficient data to assess statistical relationships or because the nature of causal order could not be confidently inferred. For example, in a hypothetical study of the determinants of government policy, in which the LNA confirmed a hypothesized relationship between institutional form and policy outcome, the SNA would likely investigate the specific actions of groups and/or individuals within a given country. This would be done in an attempt to find specific evidence that the patterns of human organization hypothesized to have been influenced by the institutional form were actually manifest in reality. Moreover, the SNA is particularly useful for investigating the impact of rival explanations for which we lack good cross-country data.

The synergistic qualities of LNA and SNA reflect the different types of data that each brings to the analysis of a problem, and their relative strengths in the task of causal inference. Here it is extremely useful to highlight the distinction between a "data-set observation," which corresponds to a row in a rectangular dataset, and a "causal-process observation," which is "the foundation for process-oriented causal inference. (It) provides information about mechanism and context" (Collier, Brady, and Seawright 2004, 253). We can say that LNA is, by definition, comprised only of dataset observations, whereas the hallmark of SNA is a much smaller number of dataset observations *and* a host of causal process observations.<sup>17</sup> Within-case analysis generally entails the scrutiny of a heterogeneous set of materials,

almost by necessity, involves less methodological structure than LNA because the analysis is strongly oriented toward discovery of novel social and political processes that take place in distinctly different ways across time and space. In recent years, there has been increasing methodological attention to the different types of strategies used by scholars when studying one or a few cases intensively. However, echoing the statements made previously with respect to LNA, this is not the place to review all of the distinctions about how such work is carried out. See, for example, contributions in Mahoney and Rueschemeyer 2003, Brady and Collier 2004, and George and Bennett (2005).

<sup>12</sup> Thanks to Phil Shively for highlighting this central point.

<sup>13</sup> See Gerring (2004) for a discussion of within-unit analysis in case studies.

<sup>14</sup> The number of rows in a dataset is typically understood as the number of country cases, or "N," that distinguishes small-N and large-N research. By now, most methodologists agree that a small-N study will have many observations, but as Collier, Brady, and Seawright (2004) point out, different inferential strategies are used to

<sup>12</sup> SNA involves multiple "within-case" observations, across space, time, and/or other dimensions. LNA may also involve multiple observations of country cases when cross-sectional data are pooled across time.

<sup>13</sup> See, for example, Ragin (1987, 69–84) for an excellent analysis of several combined approaches.

<sup>14</sup> Just as there are many styles and strategies for statistical analysis, there are at least as many approaches to SNA—an approach that,

including printed documents, interviews, and other observations that provide important information about the social phenomena we seek to understand. Because such materials are produced in such different shapes and forms across time and space, it is often impossible to specify, *a priori*, a set of very precise coding rules that would allow for an easily repeatable data collection and analysis process. These materials provide more fine-grained measurements of a host of events and behaviors, at both the micro- and macrolevels, and often in close temporal proximity to one another. Such data are virtually impossible to capture across large numbers of countries in a consistent manner. Scholars gain analytic leverage when they scrutinize the theoretical implications of these observations, either by testing existing hypotheses or by inductively developing new propositions about general relationships between causes and effects.

Although the distinction between LNA and SNA is generally between quantitative and qualitative modes of analysis, some aspects of SNA may involve quantitative analyses at *different levels of analysis*. For example, one could analyze a survey of individuals for a given country if that analysis could shed light on the dynamics of the social or political process being studied for the country at large. Analyses of individual behavior are specifically relevant to the nested approach only to the extent that they shed light on the larger questions being considered in the LNA. In a similar manner, the SNA might include time-series analysis (using country-year as the unit of observation) as a way of linking cause to effect or for dealing with case-specific rival explanations, particularly when the LNA was carried out as cross-sectional analysis. For example, in Lieberman (2003), time-series analyses of the production of government tax collections in the SNA of South Africa helped to rule out the rival explanatory power of the role of early reliance on mining revenues, which would not have been possible in the cross-country LNA, for which comparable data were not available.

The inclusion of additional theoretically valid cases is always preferred in LNA, but practical constraints on investigator skills and time as well as the desirability and feasibility of reporting in-depth analyses on multiple cases create important tradeoffs which must be weighed by scholars when selecting cases for the SNA. There is no theoretical benchmark akin to probability theory that small-N scholars can draw on to establish precise guidelines about what constitutes compelling evidence. The very nature of "causal process observations" is that they are highly heterogeneous: some documented observations may serve as particularly

powerful "smoking gun" evidence linking cause to effect, whereas others may simply serve as incremental steps that increase the plausibility of a set of theoretical claims. Small-N analysis provides the opportunity to implement various "quasi-experimental" explorations by looking at the impact of various shocks or treatments within the historical record.<sup>18</sup>

Particularly if one were to follow the recommendations of King, Keohane, and Verba (1994) to increase the number of observations, scholars might incorrectly conclude that the best strategy for the SNA component of the nested analysis would be to analyze as many country cases as possible. On the contrary, such a strategy tends to lead to a diminution of the core strengths of the SNA. Increased degrees of freedom are provided by the LNA, and nested analysis should rely on the SNA component to provide more depth than breadth—that is, given a fixed amount of scholarly resources, more energy ought to be devoted to identifying and analyzing causal process observations within cases, rather than to providing thinner insights about more cases. Because the inherent weakness of SNA is its inability to assess external validity, there is no point in trying to force it do this when the LNA component of the research design can do that work. Notwithstanding this advice, it will almost always be useful to evaluate more than one case in the SNA; the elaboration of concepts and mechanisms can best be accomplished through *comparison*. A great strength of small-N analysis is the juxtaposition of both similar and contrasting cases, helping to make transparent the operationalization of concepts that are largely hidden in the analysis of a statistical dataset. Furthermore, comparison provides an empirical basis for making narrative assessments of counter-factual claims—that is, an event would have happened a different way had the score on a key variable or set of variables been different (George 1979).

To the extent that scholars increasingly employ variants of nested analysis, standards will need to be established as to what constitutes an actual "case" study. For example, in studies that report statistical and case study findings, Reiter and Stam (2002), and Huth (1996) deploy what can be described as "mini-case analyses." These help to alert readers to examples of the argument being made by highlighting how well-known cases fit within their typologies and the degree to which they confirm to theoretical expectations. However, in these examples, the use of SNA is rather limited, and so little additional analytic value is gained. In these studies, the case analyses provide proper names for the independent and dependent variable scores, but they do not provide much elaboration about the alternative ways in which these scores were measured in comparison to the measurement procedures followed in the large-N dataset. Moreover, Reiter and Stam and Huth do not proceed with process tracing, linking cause to effect with any significant narrative. Just as statistical analyses

interpret such data. It is worth noting that even with these additional observations, such research is dubbed small-N—a convention that I use here. Meanwhile, the proliferation of TSCS analyses of country-level data is widely touted as useful strategies for increasing analytic power through a larger N (e.g., Beck and Katz 1995), but as Western and Jackman (1994, 414–5) observe, the time-invariant quality of many variables considered in cross-country analyses often implies that TSCS adds minimal additional analytic leverage for the overall problem being studied.

<sup>18</sup> See Campbell and Stanley 1966. I develop a set of strategies for exploring institutional hypotheses in small-N cross-country research in Lieberman 2001a.

must report on the sample size of the dataset, SNA demands full and clear exposition of the array of sources consulted and the depth of the historical analysis considered prior to writing the narrative.<sup>19</sup> As the number of cases in the SNA increases, the individual case analyses are likely to become increasingly superficial, and the distinct advantages of SNA are likely to diminish.

Beyond emphasizing the general complementarities, it is also important to focus the SNA based on the specific findings and analysis of the LNA. Recalling the question posed at the end of the previous section—namely, the analyst's assessment of the robustness of the preliminary LNA—SNA will then proceed along one of two tracks. If the answer is “yes, the results were robust,” as indicated in Figure 1, then the goal of the SNA will be almost exclusively focused on *testing* the model estimated in the LNA. On the other hand, if the findings were not deemed to be robust, or if one or more important hypotheses could not be explored, including if the analyst believes that the appropriate theoretical model has not yet been specified, the SNA will be oriented toward model *building*. As I detail in the sections that follow, the decision about whether to proceed with a model-testing Small-N Analysis (Mt-SNA) or a model-building Small-N Analysis (Mb-SNA) will inform the scope of the analysis, the case selection strategy, and the analysis-ending criteria for the SNA. Practitioners may respond that SNA is itself a mix of model building and model testing and that the dichotomy is a false one. Although it is true that these may be “ideal-type” approaches, there is enormous benefit to being self-conscious about the *central* intention of one's research in the SNA stage, particularly because the nested approach provides distinct sets of guidelines for the respective strategies. Assessment of the preliminary LNA constitutes an important decision-point in how the nested approach will be carried out (as depicted in Figure 1), providing important guidelines for an appropriate analytic scope for the SNA.

### Model-Testing SNA (Mt-SNA)

When scholars decide they are content with both the specification and fit of the model specified in the LNA, the main goal of the in-depth component of the nested research design is to further test the robustness of those findings. Given the potential for problems of endogeneity and poor data in statistical analyses carried out at the country level of analysis, statistical results alone rarely provide sufficient evidence of the robustness of a theoretical model. Almost inevitably, strong questions arise about causal order, heterogeneity of cases, and the quality of measurement. SNA provides an important opportunity to counter such charges. As Achen and Snidal (1989, 168–69) point out in an article otherwise quite critical of how such work is often practiced, “Case

studies are an important complement to both theory-building and statistical investigations . . . they allow a close examination of historical sequences in the search for causal processes . . . Comparison of historical cases to theoretical predictions provides a sense of whether the theoretical story is compelling.”

As the goal is to complement the LNA, the use of SNA in nested analysis should aim to gain contextually based evidence that a particular causal model or theory actually “worked” in the manner specified by the model. Can the start, end, and *intermediate* steps of the model be used to explain the behavior of real-world actors? Although this recommendation runs counter to the admonitions of Przeworski and Teune (1970), who argue that the ultimate goal should be to eliminate such labels, I believe that the nested analysis approach resonates more broadly with the general goals and expectations of scholars engaged in comparative research. That is, not only are we interested in our ability to make sense of patterns of variation, but also we would also like to use theory to account for decidedly important and seemingly anomalous outcomes in specific times and places. Moreover, unlike in some forms of medical research, where researchers are more likely to be content to find that a cause (say a drug used for minor pain relief) is related to a particular effect (say, better coronary health), even if they cannot identify the causal pathway of this relationship, social scientists are much less likely to be content with analogous findings. A good social science theory should not merely predict a particular relationship between independent and dependent variables, but it should also explain how and why these factors are related to one another (Gerring 2005), suggesting implications for what types of events and/or processes lie between cause and effect. SNA aims to make specific observations between those two points, verifying the plausibility of the stated mechanisms in terms of actions, outcomes, and/or perceptions. The SNA ought to demonstrate within the logic of a compelling narrative that in the *absence* of a particular cause, it would have been difficult to imagine the observed outcome.

In the case of Mt-SNA, scholars can justifiably focus their investigative resources on researching and analyzing the statistically significant results. The combination of theory and statistical results compels us to gather evidence—in the form of primary and secondary printed sources, interviews, surveys, and the other types of materials typically consulted for the development of an in-depth case analysis—that allows us to write a detailed narrative from the vantage-point of the preferred model. The evidence required for the SNA depends upon the nature of the theory. For instance, in a highly structural argument, actors may not be very aware of the circumstances that shape their actions, and so evidence of large-scale processes and events will be more appropriate than in the case of agent-oriented models, in which we would expect evidence of individual-level calculations and deliberate action.

While retaining a focus on assessing the plausibility of the preferred model, Mt-SNA should also aim to

<sup>19</sup> We should not establish as a standard for SNA that a longer narrative necessarily implies more careful research and/or analysis. Our assessment of the findings should be based on the methods used to gather and to analyze such data.



address two types of rival explanations.<sup>20</sup> First, if there were strong hypotheses that could not be considered in the LNA because of lack of cross-country data, the analyst should try to assess the strength of the hypothesis in the case study or studies. If cause and effect do not co-vary in the predicted manner, and/or if it is not possible to develop a coherent causal narrative guided by the rival model, the rival hypothesis can be dismissed. Second, the scholar should verify that the cause preceded the effect. Cross-country statistical databases (used in the LNA) are often highly limited in terms of temporal scope, and SNA can be used to verify that prior historical factors did not produce the observed result.

### Model-Building SNA (Mb-SNA)

When the state of theory is initially weak or refuted by the LNA and/or the quality of the cross-country statistical data is not sufficient to adequately assess the chief hypotheses, the SNA will be called on to do more work. In this instance, the nested analysis approach demands a more wide-ranging and inductive Mb-SNA. Although scholars may initiate a research project with only general theoretical hunches, Mb-SNA involves using various case materials to develop well-specified theoretical accounts of cross-country variation on the outcome of interest. Moreover, the Mb-SNA ought to be used to identify measures that are valid and reliable indicators of the analytic constructs within the theoretical model. A clear shortcoming of LNA as it is often practiced in cross-country research is that many "off-the-shelf" datasets tend to be used, and variables may not actually measure what the theory describes.<sup>21</sup> Particularly in the instance of Mb-SNA, the investigator's proximity to a wide range of data sources should facilitate the development of valid measures.

As stated at the outset, many scholars may eschew the goal of identifying broadly generalizable theories or covering laws<sup>22</sup> and may use the LNA portion of the nested analysis approach simply to point out the limits of existing data and theory, motivating a more inductive search for explanations within a single case or small set of cases. Others seek more nomothetic findings. In either case, the scholar engaged in Mb-SNA does *not* proceed with the notion that a fully specified model is available and must develop explanations for the puzzle of varied outcomes. Although the Mb-SNA approach assumes that the refuted alternative hypotheses were

adequately tested, the Mb-SNA approach invites re-examination of all theoretically strong propositions to the extent that data are available.

Inevitably, close-range analysis of one or a few country cases entails making difficult choices about which materials to investigate and which leads to pursue. Nevertheless, in most instances Mb-SNA has several advantages compared to SNA carried out in the absence of a preliminary LNA (i.e., a nonnested design). First, the scholar is equipped with useful, if partial, information about the strength of rival explanations and control variables. Of course, the reason for the negative results may be due to the poor quality of the data in the first place, but at the very least there is some indication about the weakness of relationships. Second, to the extent that the preliminary LNA provides a reasonable measure of the dependent variable, the Mb-SNA can focus on accounting for estimated differences between cases, or between cases and some central tendency of the population, having controlled for the effects of other influences. Third, the nested approach may induce the analyst to specify clearer concepts and models than conventional SNA, because even the anticipation of analyzing the results with statistical/quantitative tools implies the need for careful delineation of cause and effect. In this case, the SNA will be carried out with an eye toward theoretical parsimony and clarity, which are not always hallmarks of the SNA approach.

### CASE SELECTION STRATEGIES FOR SNA

Nested analysis provides a solution to many of the tensions that exist in the current state of methodological advice about case selection strategies: scholars often justify intensive case study work because of a sense that they lack sufficient data and analysis of such cases, and yet most case selection strategies require that we justify that selection at the outset based on what we think we know about a particular case or set of cases, often in relation to a broader universe of cases. Nested analysis provides some assistance for squaring this circle, by detailing some guidelines for the daunting task of case selection with respect to the findings of the preliminary LNA. These strategies are useful when a scholar enters a research project without a prior inclination to investigate a particular case(s) and/or for assessing the analytic utility of certain case selection choices when a scholar already has a predisposition toward those cases prior to carrying out the preliminary LNA. Indeed, the nested analysis approach can leverage the accumulation of case-relevant skills and background (including language skills, case familiarity, etc.) which are important assets for most qualitatively oriented scholars. There is rarely a perfect case selection strategy for SNA. Rather, there is a set of options and choices that, again, may be narrowed significantly by the LNA. Specifically, we can make informed choices about whether to select cases based on predicted and actual scores on the independent or dependent variable and whether or not to select cases randomly.

<sup>20</sup> For a fuller discussion of the use of qualitative research to address rival explanations, see Collier, Brady, and Seawright 2004.

<sup>21</sup> See, for example, Lieberman 2001b for a discussion of how cross-country taxation data may (or may not) correspond with theoretical constructs about the relationships between state and society. Ragin (2000) makes an important point that the scale of country-level indicators (e.g., GDP/capita) may not correspond to differences in the underlying construct (e.g., level of development), and conceptually sensitive cutpoints and calibrations may be required.

<sup>22</sup> For a thoughtful challenge to the notion that comparative analysis should always involve the pursuit of nomological covering laws, see Zuckerman 1997.

### Selecting Cases Relative to the Preliminary Model ("On-" or "Off-the-Line")

Perhaps no aspect of the methodological literature on case selection has left scholars engaged in small-N research more confused than the question of whether to select cases based on values on the independent or dependent variables. Particularly in the area of cross-national research, scholars have highlighted the pitfalls of selecting a case based on an extreme score on the dependent variable and attempting to infer general conclusions about the larger universe of cases (Geddes 1990). More recently, methodologists have highlighted a wider range of case selection options that will mitigate such problems, including the explicit accounting of the selection mechanism (King, Keohane, and Verba 1994, 128–37). More stridently, they recommend that scholars should select cases based on scores on the explanatory variable—a strategy that does not lead to analogous pitfalls of selection bias (King, Keohane, and Verba, 1994: 137–42)—while insisting that such cases be selected without knowledge of the dependent variable scores (142–46). Unfortunately, this solution, which attempts to replicate the inferential logic of experimental research, is largely impractical. In the first place, it assumes very strong theory, which is often not the case in cross-national research. In the second, because qualitatively oriented scholars tend to approach research questions from the perspective of trying to understand the determinants of puzzling outcomes, they are almost certain to know the scores on the outcome variable.

A second issue that comes up is whether we should investigate cases that are seemingly anomalous or cases that "prove" a more general point. Is the role of in-depth analysis to assess the value of preferred theories, to lead us to new propositions, and/or to gain better insights into cases deemed to be of intrinsic interest? The nested analysis approach provides a strong foundation for adjudicating among the competing goals and inferential logic associated with case selection strategies, asking the scholar to make decisions about case selection in the context of the assessment of the preliminary LNA, which includes an assessment of confidence in one's theoretical model.

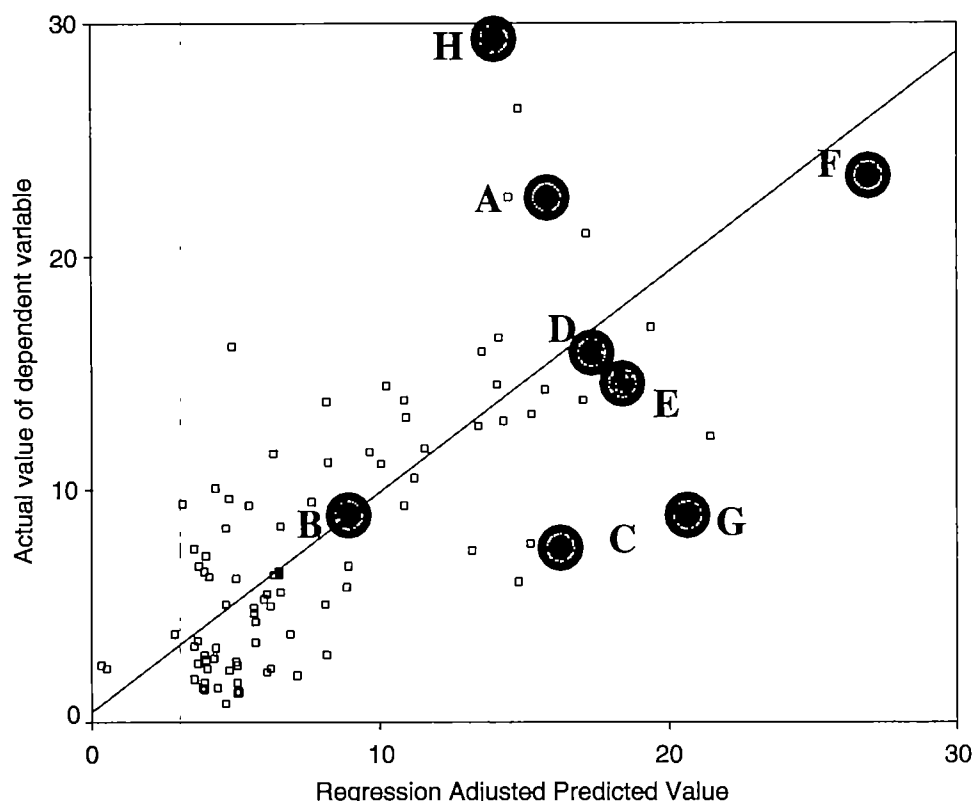
When carrying out Mt-SNA, scholars should only select cases for further investigation that are *well predicted* by the best fitting statistical model. Recall here that a decision has already been made that cases outside the confidence interval are not of theoretical interest and should be treated as unexplained "noise." Country cases that are on, or close to, the 45-degree line (plotting actual dependent variable scores against regression-predicted scores) should be identified as possible candidates for in-depth analysis. As discussed previously, in this instance SNA provides a check for spurious correlation and can help to fine-tune a theoretical argument by elaborating causal mechanisms. Although intensive investigation of "on-the-line" cases may lead to the identification of alternative explanations, the primary goal is to assess the strength of a *particular* model. As such, there is little value to the

pursuit of cases that are not well predicted by the model.

Moreover, when carrying out Mt-SNA, one should select cases based on the widest degree of variation on the *independent* or *explanatory* variables that are central to the model. Because the goal is to demonstrate the robustness of a particular causal argument, the onus on the scholar is to identify process-tracing evidence from cause to effect. The opposite approach—of starting with the outcome and working backwards—would be much less efficient given the assessment of confidence in the original model. By selecting cases with varied scores on the explanatory variables, the scholar can use the SNA to demonstrate the nature of the predicted causal effect associated with the model in contrasting contexts.

Both Swank (2002) and Martin (1992) provide examples of book-length studies in which early chapters report statistical analyses that pave the way for Mt-SNA. According to Martin (1992, 92), "For those variables that showed statistically significant effects, the analyses complement the case studies by improving our confidence in the generalizability of our results." In each case, LNA provides initial confirmations of the author's core hypotheses and dismisses several rival hypotheses. However, in both cases, the authors acknowledge that questions about causality arise and that a range of possible mechanisms could be linking independent and dependent variables. As a result, they both select cases based on different scores on the central hypothesized explanatory variables and demonstrate the plausibility of their hypotheses by tracing the impact of alternative scores on those variables to predicted outcomes in the respective cases. (Graphically, this would be akin to selecting cases such as B, D, E, and F from Figure 2, in which a range of predicted values are considered.) Both scholars are deliberate in this approach. For instance, Martin (1992, 11) writes, "This quantitative work allows me to further refine these hypotheses and provide a framework for the case studies that follow." Both Swank and Martin report additional findings and nuances about the cases they describe beyond demonstrating the plausibility of hypothesized relationships from the statistical results. For example, Swank points out that large-scale variables such as "international capital mobility" (captured in the LNA) are connected to discrete policy outcomes such as social expenditure through specific historical *episodes* (presumably distinct from an argument in which the mechanism is through long-term trends, or slow shifts), such as German unification or Italian political system restructuring (Swank 2002, 278). Within the case studies, we observe how actors behave, and we are presented with a more transparent accounting of causal mechanisms. Compared to an otherwise quite similar study such as Garrett's (1998) examination of the role of partisan politics in mediating the pressures of globalization, which presents only statistical findings, Swank's uncovering of cases and mechanisms provides significant additional evidence and insight. In the absence of such SNA, we would have been left to imagine the multiple causal pathways possibly associated with the

FIGURE 2. Case Selection from a Hypothetical Regression Analysis



statistical associations, and with greater skepticism about the general robustness—that is, non-spuriousness—of the results.

On the other hand, a very different set of strategies for case selection should be adopted in the case of Mb-SNA. First, *at least one* case that has not been well predicted by the best-fitting statistical model should be selected. Although it may be useful to select additional cases that are on the best-fit line for comparative analysis, the assessment that the preliminary statistical model was not sufficiently robust or that there were not sufficient data available to test certain critical hypotheses compels the scholar to examine cases that are *not* explainable by the right-hand-side variables included in the preliminary LNA. It is important to keep in mind the distinction between cases that are not well explained by the model (say, more than 2 standard deviations from the predicted value) and truly extreme cases that are several standard deviations from any other cases (e.g., case “H” in Figure 2). In the latter instance, the extreme nature of the case placement makes it more likely that the outcome was produced by a different causal process than most of the other cases in the population (and/or that some measurement error was involved). When such a distribution of cases is presented, case selection will hinge on whether the scholar is more interested in “making sense” of that deviance, or of developing a general theory that directly accounts for greater numbers of (less extreme) country cases.

Only when the scholar has good reason to believe that a particular case is “on-the-line” for entirely spurious reasons would it be useful to select such a case for Mb-SNA. However, in such instances the heuristic value of the preliminary LNA becomes increasingly obscured, and hence, of limited value.

In contrast to the Mt-SNA, case selection in Mb-SNA involves selection of cases based on initial scores on the *dependent* variable. Because Mb-SNA proceeds with vaguer theoretical hunches, the central goal is to try to account for important patterns of variation on the outcome.<sup>23</sup> Although it is important to try to ensure that among the cases selected there is sufficient variation on the explanatory variables of greatest interest at the outset, this is of secondary concern because there is much less confidence at the outset of the SNA that such variables will be significant when the research and analysis are complete. The very nature of Mb-SNA implies that we may lack the scores on the explanatory variables of interest at the outset of the project, making it impossible to use the explanatory variables for case selection. Although the strategy of selecting on the dependent variable has been a potential pitfall for much small-N scholarship, the nested approach provides

<sup>23</sup> Certainly, much social science analysis begins with the question, “What is the effect of X?” but almost always, there is a clear Y or outcome in mind. Such instances are examples of “strong theory.” It is very rare that a scholar will start with an explanatory variable, but search inductively for an outcome to explain.

important correctives: the preliminary LNA provides a framework for selecting cases that vary widely on the variables of interest, and to the extent that the scholar hopes to draw general conclusions about the application of the resulting model, nested analysis involves the assessment of the hypothesis in subsequent LNA (discussed in the following section). Because causal inference in the nested approach does not rely solely on the small-N portion, the standard pitfalls of selection bias are less likely to lead to faulty inferences.

Using nested analysis, the preliminary LNA can be used to motivate structured comparisons for SNA, including a mix of “on-the-line” and/or “off-the-line” cases. In the simplest manifestation, when countries that would ordinarily be predicted to have similar outcomes wind up with different outcomes, perhaps on either side of the regression line, and with at least one case outside the confidence interval, scholars are presented with useful analytic puzzles that merit further examination (e.g., cases A, D, and C in Figure 2). Along these lines, the use of the nested approach could help to expand the scope of structured focused comparisons. Although there is a long tradition of deploying variants of Mill’s “method of difference” to gain analytic leverage in cross-national comparative analysis, the requirement of identifying similar cases tends to limit scholars to comparing cases within regions, forcing certain sets of comparisons to reemerge: “France/Germany,” “U.S./Canada,” “Brazil/Argentina,” and so forth. To a large extent, the underlying logic of such comparisons requires that the scholar make the implausible argument that the two or more countries are “virtually identical” in every way except on the relevant independent and dependent variables. As typically practiced (i.e., in the absence of LNA), the method virtually precludes making comparisons of countries at different levels of economic development, because that factor is assumed to have a causal influence on most outcomes of interest to Political Scientists. For example, comparisons between the United States and India might ordinarily be dismissed as not particularly useful because of vast differences in levels of economic development. However, within a nested analysis, one might find in the preliminary LNA that indicators of development do not hold any explanatory weight for the outcome of interest, and that colonial legacy (Anglo in both cases) and state structure (federal in both cases) are important predictors of the outcome, leading to similar point estimates and compelling a focused analysis of the two countries. Alternatively, in a strategy that approximates Mill’s method of agreement, one might select cases with *differing* regression predicted values, and attempt to explain similarities in outcomes (e.g., cases B and G). In either case, LNA can set the stage for a comparative analysis that might otherwise seem implausible. The juxtaposition of such country cases allows for the additional exploration of the role of rival hypotheses that might not have been considered in the LNA because of lack of theory or data.

The nested analysis approach provides a self-conscious strategy for what many case-oriented scholars already do in practice: begin a research problem

with an intuition that a particular case defies conventional wisdom or theorizing about a particular phenomenon, and then proceed inductively to generate explanations and theories that account for that exceptionalism. When using nested analysis, a potentially important finding of the preliminary LNA is that variables ordinarily thought to be associated with the outcome turn out to be statistically unrelated in the large sample. Alternatively, if the preliminary LNA demonstrates that the case *was* well predicted by conventional variables, this would give good reason to rethink the intuition of the case’s uniqueness. If the LNA confirms the case’s outlier status, however, this provides strong justification for intensive study.

As an example of such a move, Coppedge (2005) motivates the question of patterns of regime change over-time in Venezuela through various engagements with theory and preliminary LNA.<sup>24</sup> On the one hand, he demonstrates that on its own, a variable measuring over-time changes in level of economic development does a relatively good job of predicting democratic breakdown in that case. On the other hand, the inclusion of other factors helps to provide a better fitting model of regime outcomes *more generally* (across a large sample of approximately 4,000 country-years), and such a model does *not* predict the observed over-time changes in Venezuela’s political regime. From this perspective, the need for the case study is clear: existing wisdom on the subject could *not* account for an important political outcome, and there is room for a new hypothesis or set of hypotheses to help address this conundrum. To accomplish this, Coppedge engages in an inductive Mb-SNA. (Incidentally, it is important to note that when using pooled time-series cross-section data, the “country” is still the unit about which one tries to make inferences, but the inclusion of historical data implies an interest in accounting for dynamics or historical patterns that describe each country, in the context of time-varying parameters.)

### Selecting Cases Randomly or Deliberately

Scholars using nested analysis also face choices about whether the selection of cases should be done randomly or deliberately (nonrandomly). Again, the best strategy depends largely on the goals of the SNA and also on scholarly tastes and the scholar’s familiarity with

<sup>24</sup> My definition and label of the nested approach are somewhat different from Coppedge’s (2005). He explains, “Nested induction consists of a case study nested within a large-sample quantitative analysis. This method has three steps: 1) explaining the case of interest as much as possible using large-sample empirical estimates of the impact of general explanatory factors; 2) using the large-N estimates to pinpoint what is not well explained by the general factors (the residuals), and 3) using traditional case-study methods to propose supplementary explanations for the residuals (1).” My approach is more expansive, incorporating a wider variety of research problems and results. Moreover, I opt for the label “nested analysis” instead of his “nested induction” because I see no reason to limit this form of research necessarily to inductive theory-building. Although case analysis almost inevitably demands induction, there is no reason that this approach could not be used to examine deductively derived propositions.

and access to certain case materials. In most cases, deliberate selection will be the most appropriate strategy, but there may be specific instances when, in the course of carrying out Mt-SNA, random case selection can be used to address specific concerns about investigator bias. In any event, explicit consideration of this option forces us to reflect on potential sources of bias and measurement error in SNA, which should be considered in all aspects of the nested analysis.

Though rarely used in practice, when carrying out Mt-SNA, it may be desirable and appropriate to use a *random* case selection strategy. In a work-in-progress, James Fearon and David Laitin (2005) elect to further test their statistical model (2003) with narrative analyses of a set of randomly selected cases.<sup>25</sup> Fearon and Laitin (2005) opt for this approach, arguing that the deliberate selection of cases risks high levels of investigator bias. In particular, they say that the random selection of cases can provide an opportunity for a “fresh reading from the standard literature about a country.” Moreover, they are concerned that in-depth investigation of cases they know well will induce confirmation of theories based on the very information that was used to derive the theory in the first place. Importantly, the rationale for random selection is *not* the development of a *representative* sample, as is the case in other forms of research, including survey research. The number of cases involved is simply too small to generate a useful representation of the entire population of country cases.

There are strengths and weaknesses associated with the random case selection option. On the one hand, there is good reason to believe that this strategy *should* lead to less investigator bias—However, it is only appropriate when the model specified in the LNA provides a good fit and when the investigator is less interested in identifying new hypotheses than in assessing the degree to which the logic of the theory behind the statistical model actually resonates with causes and effects within particular case histories. If a scholar can actually apply a model to a country with which he or she had little initial familiarity, confirm the independent and dependent variable scores with new measures, and find theoretically predicted links between cause and effect, such findings would provide considerable confirmation of the robustness of the model. As Fearon and Laitin (2005) suggest, a good strategy is to stratify cases based on independent and dependent variable scores in order to ensure a wide range of variation in case scores while attempting to economize on the total number of case studies carried out.

Despite certain appeal in the reduction of bias associated with random selection, the promised benefits must be weighed against pragmatic investigator limitations. The very rationale of this strategy commits scholars to cases where they may lack the technical skills for careful readings of country data, and mostly, if not exclusively, to secondary sources that may already

be heavily biased by a particular theoretical bent.<sup>26</sup> This strategy may be particularly problematic when scholars carry out research in issue areas for which a complete secondary literature does not exist (in the case of Fearon and Laitin (2005), their focus on civil wars implies that this concern does not hold), requiring scholars to probe deeply into primary materials in order to carry out the analysis. One solution would be to enlist country experts to collaborate on country-based research generated from random selection and to ask them to adjudicate among best-fitting models (while being blind to the preferred model). This is an ideal strategy from a methodological standpoint if such an opportunity is available and appeals to one's scholarly style, but it also imposes high research costs.

Although the random selection approach is an intriguing option, most scholars will likely opt for a deliberate, or nonrandom, approach to the selection of cases. Particularly because problems of selection bias do not apply in the LNA component of the nested analysis research design, minimization of this bias in the SNA component is not likely to justify the costs associated with random case selection. Indeed, as stated at the outset, many scholars are interested to see whether general theories can help to make sense of particular cases and do not view case analysis as merely a means for assessing general theories. When selecting cases deliberately, the standard benefits of SNA are much more likely to apply, including the ability of the scholar to gain access to (often highly heterogeneous) data and to sensitively analyze such data with an appropriate degree of contextual background to make valid comparisons across cases. For example, evidence of the harsh exchange of words in various legislative contexts is likely to have very different implications for how we interpret the degree of cohesion or polarization across polities, depending on the norms of parliamentary debate. Or, in the case of racial/ethnic politics, the “coding” of bigoted language and the subtle ways in which discriminatory practices get carried out may only be apparent to a seasoned analyst. Valuable field research, the quality of which is greatly enhanced through language skills, is more likely to be endeavored if country cases are deliberately selected.

Indeed, when engaged in Mb-SNA, random selection of cases should absolutely be avoided because such an approach would be tantamount to saying, “I don't have a good theory, and I don't have an intuition about why a particular case would be illuminating for constructing a theory,” which is hardly a solid foundation for investigation. Of course, many scholars who find themselves engaged in Mb-SNA will arrive at this form of analysis because, as discussed in the previously cited examples, they had already identified cases of potential theoretical interest. Alternatively, such cases will be selected because a scholar believes that he or she has a particular expertise, such as language skills,

<sup>25</sup> However, they do not limit themselves to the selection of well-predicted cases.

<sup>26</sup> For example, see Lustick (1996) for a discussion of the problems of bias in secondary sources in political science research. For a more general discussion of the problem of random selection in SNA, see King, Keohane, and Verba (1994, 125–28).

background, or historical knowledge, or because of a particular interest in a case.

When a scholar is intent on studying a particular case or set of cases, the nested analysis approach obviates the need to make the artificial claim that the case is the best one for studying a particular research question. Rather, the approach allows the scholar to identify the particular information that he or she wants to glean from the in-depth analysis of almost any case, and then to assess the potential added value of such analysis relative to a larger body of theory and data.

Scholars engaged in Mt-SNA may also use deliberate case selection, but they should avoid using the specific case or cases that informed the initial development of the theoretical model (i.e., prior to the preliminary LNA) as the basis for *testing* the model. Such a constraint may be highly prohibitive for scholars with a wide-ranging knowledge of country cases, whose theorizing may be informed by several important cases. A next best solution would be to try to gather *new* information about the particular cases with which the analyst is more generally familiar and to attempt to “test” the LNA-verified hypotheses with such data. Alternatively, the analyst may deliberately select a case of substantive interest, but with little prior knowledge of case specifics, capturing most of the benefits of the random selection procedure.

### ASSESSING THE FINDINGS OF THE SNA: THE NEED FOR FURTHER NESTING?

Moving between SNA and LNA, when taken to the extreme, could imply an endless loop of research, with the only end in sight being the intensive analysis of every country case. Clearly, this is not a helpful vision of the nested approach, both because it is impractical and because it is likely to violate social scientific preferences for parsimony (Gerring 2001, 106–7). There is always more to be learned, but it is necessary to establish a set of criteria and procedures to conclude the analysis, leaving unanswered questions to future research. Again, just as there are no absolute answers to such a question in the cases of LNA or SNA on their own, strict guidelines cannot be established for the nested analysis approach. Nonetheless, we can establish useful assessment criteria for making decisions about when to end the analysis. Contrary to Lijphart’s original view of the possible interaction between different types of research methods in comparative political analysis—in which SNA was merely a “way station” for LNA<sup>27</sup>—in the nested analysis approach, LNA serves as a way station for SNA at least as often as the reverse. A fundamental interest in the understanding of specific country cases helps to anchor the analysis in the nested research design.

Two endpoints are clear: in the case of Mt-SNA, if one or more intensive case studies can demonstrate the validity of the theoretical model—which had already passed muster in the LNA—by plausibly linking cause

to effect in the expected manner, then the nested analysis provides ringing support for the model (**End analysis I** in Figure 1). Although we do not know the exact sequence for how the analysis was actually carried out in these works, the Martin (1992) and Swank (2002) books appear to be examples of this route.

At the other extreme, in the case of Mb-SNA, if a coherent theoretical explanation for the outcomes cannot be formulated, this also implies a natural endpoint (**End analysis IV** in Figure 1). In this situation, neither LNA nor SNA could generate a robust finding, suggesting that either the research question was poorly formulated or the outcome is generated by a largely random process. This implies the project should be abandoned or substantially reconstituted to the extent that it would be recognized as a new project. In a discipline that tends not to value negative findings or atheoretical analyses, it should come as no surprise that there are no published examples of such a project.<sup>28</sup>

### When the Model-Testing SNA Fits Poorly

In between these two extremes, as depicted in Figure 1, there are a series of assessments that must be made to establish the next steps and procedures for analysis. When engaged in Mt-SNA, if the analysis does *not* support the statistical model, the scholar must assess the reason(s) for this poor fit. As in social science more generally, assessments of the link between evidence and theory contain a subjective element, and scholars are likely to disagree about goodness of fit and the factors driving the distribution of the data. Although the nested analysis approach cannot completely resolve such debates, it specifies the parameters of the assessment and the steps that ought to follow particular conclusions drawn from the data and analysis.

**Idiosyncratic Cases.** On the one hand, the scholar may decide that the Mt-SNA did not support the model because the selected case was clearly idiosyncratic in at least one important way—that is, some extremely rare historical event or set of circumstances obfuscated the types of social and political processes that were in the original model, or the variable scores were incorrectly measured for some highly anomalous, case-specific reason. Moreover, the scholar may decide that such unique circumstances do not merit theoretical elaboration because the epiphenomenal sequence of events is unlikely to travel to other cases. In this instance, the scholar remains confident that the model estimated in the LNA is still a robust one and that the case selected for in-depth SNA was found to be “on the line,” but not for the reasons justified in the model. Although it would be important to report the findings of the SNA in the analysis, the degree of emphasis on that narrative will be a question of scholarly tastes—specifically, a taste for highlighting typical cases versus puzzling or deviant cases. Nonetheless, if the poor fit is due to factors not

<sup>27</sup> As analyzed by Collier (1991, 13).

<sup>28</sup> Often referred to as the “file drawer” problem, in which our exposure to the full range of evidence is constrained by vast quantities of unpublished, and therefore inaccessible, negative results



likely to be found in the larger sample, an additional case or set of cases should be selected for additional Mt-SNA. If additional analysis again fails to confirm the original statistical model, the scholar should become highly suspicious of the assessment of idiosyncrasy and consider that the model may not be accurately capturing the general process it purports to explain. In subsequent iterations the scholar might conclude that the SNA undermined the robust findings of the LNA.

**Theoretical Flaws.** Alternatively, the Mt-SNA may reveal important shortcomings in the initial model and/or the statistical results. In such an instance, the Mt-SNA would reveal that the statistical correlation was in some way spurious—the variables are not measuring what they purport to measure, or it becomes clear that the presumed causal order of the original model is not in evidence in actual case analyses, or other variables not identified in the LNA specification are observed to be doing the causal work. For example, suppose an initial theoretical model claims that presidential systems of government lead to personalistic styles of politics, and this is somehow confirmed by the LNA. If the Mt-SNA shows clearly that cases of presidentialism tended to have personalistic political styles even prior to the introduction of democratic politics, we would have good reason to challenge the original model. What started as Mt-SNA would need to become Mb-SNA. Additional inductive exploration, combined with appeal to a broader set of theoretical propositions, is clearly necessary.

### When the Mb-SNA Suggests a New Model

Looking at the Mb-SNA side of Figure 1, an additional assessment is also required when the Mb-SNA generates a promising theoretical model. Having completed intensive study of one or a few country cases, the intensive case analysis component of the nested analysis is complete, and the only remaining assessment to make is whether the model can generate testable propositions through additional LNA.

On the one hand, if the new model relies on explanatory variables that are difficult to measure across many cases (e.g., complex cultural, institutional, or historical variables), it may not be possible to develop quantifiable indicators or a statistical estimator that captures the theoretical relationships. Or, a scholar may decide that he or she has uncovered an important theoretical anomaly that is worth explaining, but for which further LNA would provide no added value because no additional cases would score in the same way, meaning no further testing of the hypothesis could be carried out. Finally, a scholar may decide that the purpose of his or her scholarship is to use theory to understand the puzzle of a case, rather than the reverse. In any of these instances, the scholar can report the findings of the preliminary LNA and end with the SNA. In the case of Coppedge's (2005) study of democratic breakdown in Venezuela, this is clearly the path that is chosen, reflected in **endpoint III**. Coppedge is able to *explain* the specific outcome in Venezuela by highlighting that

other theoretical models cannot account for the specific deviations of this case and by identifying a unique set of causal conditions that are not captured in other theoretic accounts. He leaves it for future research to determine whether the features identified as deterministic in Venezuela can be integrated into a more general theoretical model.

Alternatively, if it is possible and desirable to develop measures of the new variable(s) and to deploy reasonable statistical tests of the model, then a Mt-LNA is clearly in order. Not surprisingly, the findings from SNA can form the basis for valid LNA. Close-range analysis of one or a few cases can be akin to developing a survey instrument through open-ended interviews and focus groups using a small sample of cases before fielding a large-scale survey. That is to say that a scholar can evaluate and/or develop indicators to be used for the measurement of a large number of cases through close-range measurement of one or a few cases.<sup>29</sup> The scholar may build on the rectangular dataset used for the preliminary LNA and add variables or create new measures for old variables. Depending on the new insights derived about the scope conditions for the model—that is, the bounding of the population of cases to which the model ought to apply—the scholar may add cases and/or remove cases from the LNA. The SNA may suggest important, theoretically informed control variables and interaction effects when close-range study highlights the implausibility of a simple linear model applying across all country cases. Finally, the scholar may test new model specifications derived from the SNA within the LNA.<sup>30</sup> Regardless of the findings, the completion of this LNA should be reported, ending the nested analysis at **endpoint II**.

An excellent example of the move from SNA to LNA is presented in Lynch's (2002) study of the age-orientation of the welfare state in the advanced industrialized countries. She derives a set of hypotheses about why some countries seem to favor older citizens through intensive study of three policy areas in Italy and the Netherlands. These are further explored in a pooled time-series cross-sectional (TSCS) analysis of social spending in 20 Organization for Economic Cooperation and Development (OECD) countries between 1960 and 1996. She is able to address the conventional wisdom generated from the welfare-state literature, ruling out several key rival hypotheses, from both a cross-sectional and a longitudinal perspective—though she points out that there are some heroic assumptions involved in the analysis of cross-national TSCS data. The statistical analysis also confirms the relationships between her independent variables (program structure and mode of political competition) and an expenditure measure of her dependent variable. Unlike other

<sup>29</sup> For discussions of the relationship between alternative measurement approaches and issues of measurement validity, see Coppedge (1999) and Adcock and Collier (2001).

<sup>30</sup> Scholars should report findings based on the entire sample as well as on the sample with the cases from the SNA removed from the sample in order to assess the degree to which the cases that were used to build the new model may be driving the results in the Mt-LNA.

studies in which LNA preceded SNA, Lynch's study is a clear example of SNA driving hypotheses and statistical tests for the LNA. Indeed, it is much easier to interpret the results of the LNA having read the intensive case analyses because one can understand how the results reflect on the machinations of politics and policy outcomes in the two countries of interest. In particular, Lynch's arguments about the central determinants of policy development were motivated by close-range study, and it is hard to imagine that such hypotheses would have been generated in the absence of such analysis. LNA allowed her to examine the extent to which such findings were unique to her initial two cases, or relevant to a wider group of countries.

As another example, Martin (1992) considers a new set of regression analyses after presenting her case study of the Falkland Islands conflict. She realizes that a potentially unique factor—the impact of military involvement on sanctions cooperation—needs to be explored. Having been convinced that military involvement affected this particular case, she moves back to the LNA, but finds that military involvement had only a negligible effect on sanctions in the larger sample (153–6). In this way, SNA helps to *motivate* the exploration of rival explanations within particular cases, and more generally. Similarly, in his comparative study of the politics of taxation, Lieberman (2003) raises the possibility that Brazil's Catholic heritage had been a determining factor in the development of a tax state that was very different from South Africa's. Although the SNA uncovered no plausible evidence linking Brazilian taxation to this legacy, this rival hypothesis could be dismissed with additional confidence through further LNA which provided no statistical support for the alternative hypothesis.

Again, it is important to emphasize that the nested analysis approach presumes interest in positive and negative findings, and in the analysis of general patterns and of specific cases. If the Mt-LNA is robust, we have arrived at findings quite similar to those of endpoint I: two sets of empirical analyses confirm the validity of the results and the scholar can feel extremely confident in the general applicability of those results. If the Mt-LNA is not robust—if the new variables do not predict what we had hypothesized or if the larger model falls apart, then the scholar is left to explain why those results might not have applied in the LNA. It is up to the scholar to account for the more limited scope of the explanation, which needs to be understood in the context of the larger population of cases. Future research projects may be used to develop models with more general applicability, but in this instance, the scholar should report what has been discovered through the nested analysis.

## CONCLUSION

Despite the constraints of a relatively small, finite, complex, and heterogeneous universe of cases for analysis, scholars continue to be interested in questions about the causes and consequences of patterns of politics at the country level of analysis. To date, existing strate-

gies of large-N cross-country regression analysis, as well as small-N case study and comparative analysis, have each been found wanting. This article argues for a combined approach. Some scholars may already be practicing a variant of this approach in their analysis, but to a large extent, the steps involved in the analysis are not being fully reported. With the provision of a more complete specification of how this approach can work, scholars should find it easier to use the approach in a self-conscious manner and to provide a more transparent accounting of their research. This will facilitate evaluation and replication of results, provide greater analytic clarity (by demonstrating how various analytic results relate to one another), and provide a recognizable bridge between research traditions that often remain quite isolated from one another.

Nested analysis provides a stronger basis for causal inference than the sum of its small-N and large-N parts. Rather than emphasizing the *common* inferential logic of qualitative and quantitative research strategies—which is the hallmark of King, Keohane, and Verba's (1994) influential treatment of research methodology—the nested analysis approach emphasizes the complementary distinctiveness in these two modes of analysis and strategies for causal inference. The use of the mixed strategy helps to overcome potential sources of bias and to sort out spurious findings that might be produced in either SNA or LNA when carried out in isolation. The approach is particularly well suited to cross-national analysis, where investigators tend to be interested not only in general patterns (as one might be in the study of, say, voting behavior) but also in the analysis of specific country cases.

There are clearly real and perceived costs of integrating LNA and SNA. Perhaps most importantly, this seems to imply a substantial addition of work. Is this really two projects in one? Undoubtedly, more investigator effort is required than if the individual SNA or LNA components were used in isolation, because nested analysis demands multiple forms of measurement and causal inference, but it does not entail a simple addition of effort. Rather, by highlighting the specific utilities of each analytic strategy, the approach lightens the inferential burden that would ordinarily be carried by SNA or LNA when performed on their own. Moreover, the advent of the Internet and an accumulation of research continue to expand the scope of available datasets that may be usable. For example, in the area of democratization research, the Freedom House, Polity, and a host of other datasets provide time-varying indicators across a large number of countries. For students of the political economy of development, the World Bank, the OECD, and the International Monetary Fund publish time-varying economic and other data across most countries for several decades. Similarly, not all "in-depth" case analyses involve multiple years of field research. As Fearon and Laitin (2005), Swank (2002), Adserà and Boix (2002), and Reiter and Stam (2002) demonstrate, at least some of the benefits of SNA can be captured using readily available data sources without extensive primary research. Again, increasing access to a range of primary and secondary sources through the

Internet has made the research and analysis of cases and structured comparisons far easier than was the case for previous generations of scholars. Given constraints on the particular skills of any single investigator, the nested approach is well suited to collaboration. Although it is certainly possible that this approach could simply be used as a model for a more general dynamic of the research cycle, the particular strategies and tactics outlined here assume the combining of strategies and require clear consistency in the use of concepts and measures, which are often lost when different scholars respond to prior iterations of the same question.

Nested analysis is a pragmatic and methodologically defensible scheme for comparative analysis. In this article, I have detailed its potential benefits, not by merely accepting the compatibility of qualitative and quantitative modes of analysis, but by demonstrating how each can be used to inform the execution and interpretation of the other.

## REFERENCES

- Achen, Christopher H. 1982. *Interpreting and Using Regression*. Beverly Hills, CA: Sage.
- Achen, Christopher H., and Duncan Snidal. 1989. "Rational Deterrence Theory and Comparative Case Studies." *World Politics* 41 (January): 144–69.
- Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–47.
- Adserà, Alicia, and Carles Boix. 2002. "Trade, Democracy and the Size of the Public Sector: The Political Underpinnings of Openness." *International Organization* 56 (Spring): 229–62.
- Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. 1998. *Analytic Narratives*. Princeton, NJ: Princeton University Press.
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89 (September): 634–47.
- Bennett, Andrew. 2002. "Where the Model Frequently Meets the Road: Combining Statistical, Formal, and Case Study Methods." Presented at the Annual Meeting of the American Political Science Association, Boston.
- Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.
- Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.
- Collier, David. 1991. "The Comparative Method: Two Decades of Change." In *Comparative Political Dynamics*, ed. D. A. Rustow and K. P. Erickson. New York: Harper Collins, 7–31.
- Collier, David. 1999. "Building a Disciplined, Rigorous Center in Comparative Politics." *APSA Comparative Politics Section Newsletter* 10 (Summer): 1–2, 4.
- Collier, David, and James Mahoney. 1996. "Insights and Pitfalls." *World Politics* 49 (October): 56–92.
- Collier, David, Henry E. Brady, and Jason Seawright. 2004. "Source of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.
- Coppedge, Michael. 1999. "Thickening Thin Concepts and Theories." *Comparative Politics* 31 (July): 465–77.
- Coppedge, Michael. 2005. "Explaining Democratic Deterioration in Venezuela Through Nested Induction." In *The Third Wave of Democratization in Latin America*, ed. F. Hagopian & S. Mainwaring. Cambridge, UK; New York: Cambridge University Press.
- Fearon, James, and David Laitin. 2005. "Civil War Narratives." Theory and Research in Comparative Social Analysis Working Paper Series, 27. Department of Sociology, UCLA. <http://repositories.cdlib.org/uclascoc/trcsa/27>.
- Fearon, James, and David Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (February): 75–90.
- Garrett, Geoffrey. 1998. *Partisan Politics in the Global Economy*, *Cambridge Studies in Comparative Politics*. Cambridge, UK, New York: Cambridge University Press.
- Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* (Spring): 131–50.
- George, Alexander L. 1979. Case Studies and Theory Development: The Method of Structured, Focused Comparison. In *Diplomacy: New Approaches in History, Theory and Policy*, ed. P. G. Lauren. New York: The Free Press.
- George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gerring, John. 2001. *Social Science Methodology: A Critical Framework*. Cambridge, UK; New York: Cambridge University Press.
- Gerring, John. 2004. "What Is a Case Study and What Is It Good For?" *American Political Science Review* 98 (May): 341–54.
- Gerring, John. 2005. "Causation: A Unified Framework for the Social Sciences." *Journal of Theoretical Politics* 17 (April): 163–98.
- Huth, Paul K. 1996. *Standing Your Ground: Territorial Disputes and International Conflict*. Ann Arbor: University of Michigan Press.
- Jackman, Robert W. 1985. "Cross-National Statistical Research and the Study of Comparative Politics." *American Journal of Political Science* 29 (February): 161–82.
- Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.
- King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Laitin, David. 2002. "Comparative Politics: The State of the Sub-discipline." In *Political Science: The State of the Discipline*, ed. I. Katznelson and H. V. Milner. New York; Washington, DC; Norton: American Political Science Association.
- Lieberman, Evan. 2001a. "Causal Inference in Historical Institutional Analysis: A Specification of Periodization Strategies." *Comparative Political Studies* 34 (9): 1011–35.
- Lieberman, Evan. 2001b. "Taxation Data as Indicators of State-Society Relations: Possibilities and Pitfalls in Cross-National Research." *Studies in Comparative International Development* 36 (January): 89–115.
- Lieberman, Evan S. 2003. *Race and Regionalism in the Politics of Taxation in Brazil and South Africa*, *Cambridge Studies in Comparative Politics*. Cambridge, UK; New York: Cambridge University Press.
- Liphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65 (September): 682–93.
- Lustick, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90 (September): 605–18.
- Lynch, Julia. 2002. "The Age of Welfare: Citizens, Clients, and Generations in the Development of the Welfare State." Ph.D. diss. University of California, Berkeley.
- Mahoney, James, and Dietrich Rueschemeyer. 2003. *Comparative Historical Analysis in the Social Sciences*, *Cambridge Studies in Comparative Politics*. Cambridge, UK; New York: Cambridge University Press.
- Martin, Lisa L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.
- McKeown, Timothy. 2004. "Case Studies and the Statistical World-view." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.
- Munck, Gerardo L. 1998. "Canons of Research Design in Qualitative Research." *Studies in Comparative International Development* 33 (Fall): 18–45.

- Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.
- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Reiter, Dan, and Allan C. Stam. 2002. *Democracies at War*. Princeton, NJ: Princeton University Press.
- Rogowski, Ronald. 1995. "The Role of Theory and Anomaly in Social-Scientific Research." *American Political Science Review* 89 (June): 467-70.
- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *The American Political Science Review* 64 (December): 1033-53.
- Swank, Duane. 2002. *Global Capital, Political Institutions, and Policy Change in Developed Welfare States, Cambridge Studies in Comparative Politics*. Cambridge; New York: Cambridge University Press.
- Tarrow, Sidney. 1995. "Bridging the Quantitative-Qualitative Divide in Political Science." *American Political Science Review* 89 (June): 471-4.
- Western, Bruce, and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88 (June): 412-23.
- Zuckerman, Alan S. 1997. "Reformulating Explanatory Standards and Advancing Theory in Comparative Politics." In *Comparative Politics: Rationality, Culture, and Structure*. ed. M. I. Lichbach and A. S. Zuckerman. Cambridge; New York: Cambridge University Press.

# No Rest for the Democratic Peace

DAVID KINSELLA *Portland State University*

*Proponents of the democratic peace are accustomed to criticism. Early refutations of the research program's findings focused on questions of measurement and statistical inference. Skepticism about such matters has not fully subsided, but many more now accept the democratic peace as an empirical regularity. The aim of recent complaints has shifted to democratic peace theory. The typical approach has been to highlight select historical events that appear anomalous in light of the theory and the causal mechanisms it identifies. Sebastian Rosato's (2003) is one such critique, noteworthy for the range of causal propositions held up for scrutiny and the unequivocal rejection of them all. But Rosato fails to appreciate the dyadic logic central to democratic peace theory, and much of his criticism is therefore misdirected. Those cases that remain unexplained by the theory are not especially problematic for this progressively evolving research program.*

Sebastian Rosato (2003) has given us another spirited critique of the democratic peace project. His argument is similar to other realists' claims that the correlation between democratic-state interaction and peace is spurious, better understood as a function of power, threat, and national interests. His approach differs from others in that he attempts to scrutinize the many causal propositions contained in democratic peace theory, concluding in the end that all of them are contradicted by empirical evidence, and are consistently contradicted. But it fails on at least two counts. First, most of what Rosato cites as evidence against democratic peace theory does not in fact contradict the theory. Second, the evidence that does contradict the theory, in addition to being widely known among democratic peace researchers, is not particularly damaging to the theory, which continues to evolve at the core of a progressive research program.

The democratic peace is a dyadic empirical phenomenon. The empirical evidence that democracies rarely fight *each other* is robust, and most theoretical efforts have kept this finding front and center. Yet Rosato (2003, 589, 596), at various points in his critique, suggests that the dyadic claim is a retreat from some original monadic position in the face of arguments and examples to the contrary. Thus, dyadic propositions are cast as "restatements" or "new arguments" designed to "rescue" the theory's causal logic. This mischaracterizes the evolution of the democratic peace research program. Although some studies have offered evidence that democratic states generally conduct their foreign affairs more peacefully than non-democratic states (Benoit 1996; Ray 1995; Rousseau et al. 1996; Rummel 1995), the early theoretical and empirical work on the democratic peace, and most of what has followed, recognizes that a core element of democratic peace theory must be located in the nature of democratic states' *interaction*. Doyle (1983a, 1983b), one of the founders of the democratic peace project, is

very clear on this score: "liberalism is not inherently 'peace-loving'; nor is it consistently restrained or peaceful in intent." It has, however, "strengthened the prospects for a world peace established by the steady expansion of a separate peace among liberal societies" (Doyle 1983a, 206; see also Russett and Starr 1981, 439–44).

Rosato (2003) is well aware of the dyadic argument, but he does not seem to take it seriously. In dissecting the normative explanation, he identifies two links in the causal chain connecting domestic conduct in democracies to peaceful conduct in foreign affairs: elites externalize their norms of negotiation and non-violent conflict resolution, which in turn encourages them to trust and respect their counterparts in other democracies. If this is the case, Rosato believes, then democracies should have a record of fighting wars only in self-defense or to prevent egregious violations of human rights. Clearly democracies have not limited themselves to such conflicts and Rosato produces a list of wars fought for other, imperial reasons; this is supposed to refute the claim that democracies "generally externalize their internal norms of conflict resolution" (589, 590, my emphasis). The list does refute the claim, of course, but it is not a claim made by the corpus of democratic peace theory.

According to most variants of the theory, democratic restraint is conditioned on expectations about the conduct of the other party in the interaction, expectations informed by the other's internal political processes.<sup>1</sup> We need to know something about those processes (or perceptions of those processes) if the cases are to be counted as anomalies. Rosato (2003) acknowledges the rebuttal, but again does not take it seriously, insisting that "[t]he key to this logic is that democracies must *reliably* externalize democratic norms" (590, my emphasis). Ultimately, however, his assertion is much stronger than this: "[l]iberal states have *consistently*

David Kinsella is Associate Professor, Mark O. Hatfield School of Government, Portland State University, P. O. Box 751, Portland, OR 97207-0751. (kinsella@pdx.edu).

My thanks to Bruce Russett for some very helpful suggestions and comments.

<sup>1</sup> Russett and Oneal (2001, 49–52) discuss the dyadic focus of democratic peace research, but go on to suggest that more recent research may be pointing toward the conclusion that democracies generally are more peaceful than nondemocratic states, especially when considering which side in a mixed dyad initiates or escalates a militarized dispute.

violated liberal norms when deciding to go to war" (590, my emphasis). If this is not true by definition—isn't the decision to go to war, in the end, always a violation of liberal norms of conflict resolution?—then it is hard to imagine the type of evidence that would count against it. And even if democratic states did reliably externalize their norms, Rosato maintains that "[s]hared democratic values provide no guarantee that states will both trust and respect each other" (592). If it has come to making guarantees, then democratic peace theory surely must throw in the towel.

The dyadic logic of democratic peace theory is also set aside when Rosato (2003) turns to explanations focusing on the institutional constraints operating in democracies. He finds unconvincing the classical liberal argument that mass publics, because they bear the costs of war, have an interest in peace, and that mass publics in democracies, because their voices are heard, are a force for peace. Nor does he buy the variation on this argument, which states that certain groups within society, if not the masses, are advocates of peace, and their views are more likely to have an impact on the foreign policies of democracies than those of nondemocracies. That democratic publics and interest groups are not always pacific has long been established in public opinion research (Mueller 1973), and democratic leaders often look forward to a rally-'round-the-flag effect even when the balance of prewar opinion tilts against the use of force.

Rosato (2003) cites several examples of supportive (or quiescent) democratic publics during wars fought for reasons other than self-defense—but all of them involved nondemocratic opponents. Noting the character of opponents is the sort of "restatement" he dismisses as an attempt to save the theory from contradictory evidence—a charge that sticks only if one paints dyadic democratic peace theory as a retreat from the monadic argument, which it is not. Moreover, the examples adduced to falsify the claim that "democratic citizens are only averse to costs in their relations with other democracies" include colonial conflicts between Britain and France during the first half of the 19th century, when France was not democratic, and between Ecuador and Peru during the 1990s, when Peru was not democratic (596, note 16). During the 1830–32, 1838–41, and 1844 confrontations with Britain, the Polity Project locates France at –1 on their democracy–autocracy scale ranging from +10 to –10; whereas in the 1990s, Peru is scored as +1 (and –3 in 1992). Even if Rosato has some reason to believe that the regimes ought to be considered democratic, he gives us no indication of prowar public sentiments in these or any of the other democratic societies involved in the crises.<sup>2</sup> After all,

the stated purpose of his analysis is not to challenge the "powerful empirical generalization" that democracies rarely fight each other, which "remain[s] robust" (585), but to dispute the causal mechanisms that purportedly steer democracies away from war with each other.

Few would deny that hawkish interest groups often prevail in domestic debates or that "pacific interest groups may not generally influence the foreign policies of democratic states" (596). In the case of the recent Iraq War, there was indeed surprisingly little debate in the United States—until after the war. Rosato (2003) goes further, hypothesizing that, when contemplating going to war, autocratic leaders are *more* constrained by domestic constituents than are democratic leaders. He believes this may be true because wartime taxation without representation threatens to mobilize domestic opposition to nonrepresentative political institutions, sweeping away the autocracy in the process. This is an interesting argument, perhaps, as long as it applies to the avoidance of very costly wars. Autocrats do not typically shy away from taxation in pursuit of personal enrichment—presidential palaces and Swiss bank accounts—for fear of domestic disapproval, so they are unlikely to avoid foreign conflicts that they expect will not be terribly costly. In the end, the persuasiveness of Rosato's own causal logic will turn on the evidence. Curiously, although Rosato cites them to support his statement that autocracies "often represent groups that have a vested interest in avoiding foreign wars" (597), Peceny, Beer, and Sanchez-Terry (2002, 25) find "no unambiguous evidence of a dictatorial peace"; "only joint democracy was consistently related to a lower frequency of militarized disputes."<sup>3</sup>

The possibility that autocrats exercise more restraint in international crises is also raised in the discussion of political accountability. The argument found in democratic peace theory is that democratic leaders risk removal from office after unsuccessful and/or costly wars, a risk that is much diminished for autocratic leaders (Bueno de Mesquita et al. 1999, 2003, chap. 6; Gelpi and Griesdorf 2001; Reiter and Stam 2002). Rosato (2003, 594) disputes this logic, reasoning instead that a democratic leader is no more accountable than an autocratic leader "who is unlikely to lose office but

<sup>2</sup> He does refer us to some case studies, however. Disputes concerning the proper classification of regime types have characterized the debate between democratic peace researchers and their critics from the beginning. Rosato (2003, 600) asserts that "the farther we go back in history the harder it is to find a consensus among scholars and policymakers on what states qualify as democracies." That is probably true, but among quantitative researchers, both partisans of the democratic peace and skeptics, the classification scheme of choice

is the Polity Project (Marshall and Jaggers 2002). As far as I know, those who collect and maintain the Polity data are not invested one way or another in the democratic peace debate (see, e.g., Layne 1997, 65). Rosato's cited source for regime classification is Przeworski et al. (2000), who also are not participants in the debate, but their data cover the 1950–90 period only. Prior to 1950—the period covered by all of Table 1—he determines regime type himself, apparently using Przeworski et al.'s criteria. Likewise for the period after 1990. We are not told why he finds Polity's judgment to be wrong—way wrong—in the cases he cites.

<sup>3</sup> Peceny, Beer, and Sanchez-Terry (2002) show that of the various autocratic pairings, only those involving two single-party states have a reduced likelihood of militarized dispute, controlling for other factors. Their causal argument rests on these regimes' shared commitment to socialism, and thus is analogous to the normative explanation for the democratic peace. I assume Rosato (2003) would also reject the socialist norms argument as flawed causal logic.

can expect to be punished severely in the unlikely event that he is in fact removed." "Fear" is perhaps a better word for what the autocratic leader is feeling here—certainly the leader is not "answerable" in the sense understood by political theorists (e.g., Pitkin 1972, 55–9)—but Rosato's point is worth considering. To support the contention, he reports that after participation in costly wars, a larger percentage of autocrats than democrats are removed from office, and a larger percentage are punished (594, Table 4). He finds that after losing wars, democrats, not autocrats, are more likely to be removed from office (though not punished), but he dismisses this contrary result. "This evidence is not strong," he says, because there are so few instances of democratic losers. Rosato is right, but his evidence that autocrats are more likely to be removed from office as a consequence of involvement in costly wars is also weak. The relative infrequency of democratic involvement in both lost and costly wars argues against making much of these differences.<sup>4</sup>

A better interpretation of the results is that democrats tend to avoid wars they do not expect to win with modest cost. Rosato (2003, 594, note 14) rejects the plausibility of this "selection effect," but his reasoning is suspect. He refers to Desch's (2002, 23) calculations that the marginal effect of democracy on the probability of victory is lower than the marginal effects of other predictors, like terrain and military capabilities. Even if these calculations are taken at face value, they are irrelevant. The selection effects argument is not that democratic governance per se increases the likelihood of winning, but that democracies have access to better information about the likelihood of winning—whatever the factors contributing to victory—and are more inclined to stay out of conflicts when this information suggests that war is a losing proposition.<sup>5</sup> This means that militarized disputes between democracies, if they do occur, are more likely to become especially bloody affairs, and are avoided by leaders concerned with their political survival. The dyadic logic of democratic peace theory thus pertains to the probability of such nonevents, and the challenge for empirical investigation is well beyond the reach of Rosato's *ex post* evidence on office removal and punishment rates (Smith 1999). If fear of punishment is supposed to serve as a restraint on autocrats' propensity to resort to ill-conceived wars, what his evidence tells me is that a fair number of them have not gotten the message.

If there were a dictatorial or autocratic peace alongside the democratic peace, the causal logic explaining

it almost certainly would be dyadic (e.g., Peceny, Beer, and Sanchez-Terry 2002). Rosato's autocratic constraints proposition is intriguing, to say the least, but to date the empirical evidence has not shown autocracies to be generally less disputatious than other regime types. Either way, his critique of democratic peace theory stumbles on just this point. Aware of the dyadic arguments found in the literature, he nevertheless does not take the dyadic logic of the theory seriously. If democratic dyads are more than the sum of the democratic monads, as virtually all proponents of the democratic peace maintain, then the theory does not collapse under the weight of evidence suggesting less-than-virtuous behavior by democratic states.

Among the starkest empirical anomalies for democratic peace theory are those instances of American military interventions against other, weaker democratic regimes, so Rosato is correct to once again draw our attention to such cases. However, his list of seven or eight anomalies (590, Table 2) is longer than most democratic peace researchers will concede. The U.S. intervention in a democratic Chile in 1973 is beyond dispute, and in Guyana—not formally independent in 1961—American subversion occurred during a time of limited democratic self-government. Brazil was democratic in the early 1960s, but Rosato says the U.S. role in Quadros's resignation is unclear. Guatemala might be called democratic under Arbenz, but the Polity Project locates the regime only at +2 on their democracy–autocracy scale. The other three targets of American intervention are even less democratic according to Polity: Nicaragua, Indonesia, and Iran (each with a scale value of –1). In the cases of Indonesia and Iran, Rosato's own source classifies these regimes as "bureaucracies"—that is, "institutionalized dictatorships" (Przeworski et al. 2000, 32, 65).

Regardless of how these cases are ultimately judged, most proponents of the democratic peace are probably not inclined to quarrel with Rosato's conclusion that at least some of the American interventions are at odds with the normative logic of the theory. The real difference of opinion concerns the implications of these and other anomalies for the theory-building enterprise. Throughout his critique, Rosato adopts a falsificationist stance, suggesting that in the face of historical cases that belie the causal logic he distills from the democratic peace literature, the theory should be thrown out. Actually, Rosato does not devote much effort to revealing flawed *logic*.<sup>6</sup> Instead, he recites a list of *empirical* exceptions to the democratic peace—many of which are acknowledged as such by democratic peace proponents and some others that are not—while taking extra care to identify the causal mechanisms, postulated in democratic peace theory, that nevertheless seem to have gone missing in these cases. Thus, in regard to one such mechanism, he states that "whenever we find several examples of a democracy using military force against other democracies, the trust and respect

<sup>4</sup> Although Rosato is not inferring from a sample to a population, one indication that he overstates the difference between democratic and autocratic political survival rates due to costly wars is that it would fail a *t* test for statistical significance ( $t = 0.65, p = 0.53$ ).

<sup>5</sup> In addition to the selection effects explanation, Reiter and Stam (2002) also examine a warfighting explanation, which does posit that democratic governance affords certain advantages on the battlefield. Although Rosato (2003) relies on Desch (2002) to refute the selection effects argument, Desch's logic and methodology are severely flawed; see Reiter and Stam 2003 and Lake 2003.

<sup>6</sup> For an analysis of the logic of democratic peace theory, see Zinnes 2004.



mechanism, and therefore the normative logic, fails an important test" (591). Many will not agree that Rosato has refuted the dyadic hypotheses, but even accepting those particular refutations would not mean accepting that democratic peace theory itself has been falsified. The more fundamental problem is that the hypotheses Rosato derives from his rendition of democratic peace theory, and presumes to test, are too often monadic and do not square with the theory's prevailing dyadic logic.

Rosato (2003) states clearly at the outset that the democratic peace project has discovered a "powerful empirical generalization." He simply wants to replace their theory with an explanation centering on U.S. hegemony in the Americas and Western Europe, where most democracies happen to be located during the cold war period. Although elaborating his alternative "imperial peace" theory is not the main thrust of his critique, his brief presentation of the argument does suggest that, maybe, his is—to use the distinction drawn by Lakatos (1970)—a "sophisticated," as opposed to "naive," falsificationism. At various places in his essay, his complaints are directed at democratic peace theory as a degenerating research program.<sup>7</sup> Owen (1997), for instance, is taken to task for his attempt to "repair" the theory by introducing perceptions: to wit, what matters to democratic elites, when they contemplate resorting to force, is whether they *perceive* their opponents as liberal, not whether they *are* liberal. Elsewhere, he refers to "*ad hoc*" adjustments and other attempts to "rescue" the theory's logic (589–90, 596).

Scrutinizing research programs for signs that they may be degenerating is essential for scientific progress, but Waltz (1997) makes a useful point about the difference between theory and the application of theory as the target of scrutiny. In response to Vasquez's (1997) critique of neorealism as a degenerating research program, Waltz argues that although the concept of "threat" is introduced by Walt (1987) for purposes of applying balance-of-power theory to some seemingly anomalous cases, it does not thereby become part of the theory. More generally, there does appear to be a strong temptation to call on perceptions—perceptions of intentions in the case of Walt, perceptions of liberalism in the case of Owen (1997)—when the application of theory confronts discordant diplomatic behavior. Rosato is right to say that we are "unlikely to be able to predict how democracies will classify other states' regime type with a high level of confidence" (592); the temptation to revise theory ought to be resisted.

However, the attempt to explain anomalies by looking more closely at the perceptions of the actors involved is a worthy endeavor, as it improves our understanding of particular events. This sort of analysis may suggest that a revision of theory is in order if, for example, actors' perceptions are shown to be systematically biased under certain conditions, but it need not. And the undertaking of such studies is not perforce an indication that a research program is degenerating.<sup>8</sup>

There is a curious omission from Rosato's (2003) wide-ranging critique. Although he is aware of their analysis, the game-theoretic model of the democratic peace developed by Bueno de Mesquita et al. (1999) does not receive the attention it deserves in Rosato's discussion of political accountability (593–94). The omission is curious because Bueno de Mesquita and his colleagues offer a logically coherent theory that explains not only the propensity of democracies to remain at peace with each other but also many (I think most) of the empirical anomalies that Rosato finds problematic for democratic peace theory: namely, that democracies have often fought wars for reasons other than self-defense, including colonial wars; and that democracies have often attacked or destabilized weaker, nonthreatening states, including other democracies. Their model abandons the normative logic of democratic peace theory and retains just one basic element of the institutional logic—that a democratic government depends, for its political survival, on a larger constituency (winning coalition) than does a nondemocratic government. Beyond that, all the model assumes is that political leaders do in fact want to stay in power, and the policies they pursue, which yield a mix of public and private goods, are directed toward that end. It is thus in keeping with the democratic peace research program by virtue of the centrality of regime type in the theory.<sup>9</sup>

Whether Rosato's (2003) "imperial peace" theory represents a progressive problemshift—again, the term is Lakatos's (1970)—relative to this or other constructive efforts within the democratic peace project remains to be seen.<sup>10</sup> Its focus on American hegemonic

<sup>8</sup> The fact remains that researchers who do focus on the role of perceptions as an auxiliary factor in explaining the democratic peace often feel compelled to interpret their findings as calling for a revision of democratic peace theory. Thus, Owen (1997, 15) believes that "if liberal peace is real, a theory is needed to account for these perceptions." Rosato's (2003) frustration is understandable.

<sup>9</sup> The key intuition is that the political survival of democratic elites is relatively more dependent on the distribution of public goods, whereas the political survival of autocratic elites is more easily assured by the distribution of private goods. Because public goods are made available by successful public policies (including foreign policies), democratic leaders devote more resources to policy success, especially success in war. Democratic leaders, knowing that their democratic counterparts also try hard to succeed, avoid military confrontations with them, but not with their autocratic counterparts. Nor do they avoid confrontations with significantly weaker states, including democracies, because regardless of those states' level of effort, it is not likely to affect the outcome. The model is more fully developed and tested in Bueno de Mesquita et al. (2003).

<sup>10</sup> For an extended discussion of the applicability of Lakatos's (1970) criteria for appraising scientific progress in international studies, see Elman and Elman (2002). Chernoff (2004) provides a favorable

<sup>7</sup> "Falsification' in the sense of naive falsificationism (corroborated counterevidence) is not a *sufficient* condition for eliminating a specific theory: in spite of hundreds of known anomalies we do not regard it as falsified (that is, eliminated) until we have a better one" (Lakatos 1970, 121). Of course, when it comes to the democratic peace, not even the most committed proponents would tolerate "hundreds of known anomalies." Still, Lakatos's stipulation regarding the availability of a better theory is clear. That Rosato (2003) seemingly accepts the sophisticated falsificationist position is my interpretation of his critique, he is not explicit about his philosophical stance regarding the cumulation of knowledge in international relations and does not use the term "degenerating research program."

power as the key explanatory factor will displease most outside the realist tradition. Be that as it may, that Rosato prefaces the brief summary of his theory by restricting its temporal and spatial scope—that is, to the post-World War II period and to the Western Hemisphere and Western Europe—is not promising.<sup>11</sup> Neither is his blanket dismissal of every causal argument contained in an alternative theory that has nevertheless received extraordinarily robust empirical support by social science standards. Parsimony may be an admirable quality of realist international relations theory, but we should be wary of essentially mono-causal explanations put forward with such conviction. A virtue of the democratic peace research program has been a willingness to represent competing arguments in their multivariate models—including realist hypotheses, like Rosato's, that regional hegemony has a pacifying effect on conflict propensity. Indeed, empirical researchers working in this tradition have done much to confirm the validity of certain realist propositions, even while demonstrating the limits of realist theory. Nevertheless, there seems to be no rest for the democratic peace.

## REFERENCES

- Benoit, Kenneth. 1996. "Democracies Really Are More Pacific (in General)." *Journal of Conflict Resolution* 40 (December): 636–57.
- Bueno de Mesquita, Bruce, James Morrow, Randolph Siverson, and Alastair Smith. 1999. "An Institutional Explanation of the Democratic Peace." *American Political Science Review* 93 (August): 791–807.
- Bueno de Mesquita, Bruce, Alastair Smith, Randolph Siverson, and James Morrow. 2003. *The Logic of Political Survival*. Cambridge, MA: MIT Press.
- Chernoff, Fred. 2004. "The Study of Democratic Peace and Progress in International Relations." *International Studies Review* 6 (March): 49–77.
- Desch, Michael C. 2002. "Democracy and Victory: Why Regime Type Hardly Matters." *International Security* 27 (Fall): 5–47.
- Doyle, Michael W. 1983a. "Kant, Liberal Legacies, and Foreign Affairs." *Philosophy and Public Affairs* 12 (Summer): 205–35.
- Doyle, Michael W. 1983b. "Kant, Liberal Legacies, and Foreign Affairs, Part 2." *Philosophy and Public Affairs* 12 (Fall): 323–53.
- Elman, Colin and Miriam Fendius Elman. 2002. "How Not to Be Lakatos Intolerant: Appraising Progress in IR Research." *International Studies Quarterly* 46 (June): 231–62.
- Farber, Henry S., and Joanne Gowa. 1997. "Common Interests or Common Politics? Reinterpreting the Democratic Peace." *Journal of Politics* 59 (May): 393–417.
- Gelpi, Christopher S., and Michael Griesdorf. 2001. "Winners or Losers? Democracies in International Crises, 1918–94." *American Political Science Review* 95 (September): 633–47.
- Gowa, Joanne. 1999. *Ballots and Bullets: The Elusive Democratic Peace*. Princeton, NJ: Princeton University Press.
- Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, eds. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.
- Lake, David A. 2003. "Fair Fights? Evaluating Theories of Democracy and Victory." *International Security* 28 (Summer): 154–67.
- Layne, Christopher. 1997. "Lord Palmerston and the Triumph of Realism: Anglo-French Relations, 1830–48." In *Paths to Peace: Is Democracy the Answer?*, ed. Miriam Fendius Elman. Cambridge, MA: MIT Press.
- Marshall, Monty G., and Keith Jagers. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2002." University of Maryland. Typescript.
- Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: Wiley.
- Oneal, John R., and Bruce Russett. 1999. "The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organizations, 1885–1992." *World Politics* 52 (October): 1–37.
- Owen, John M. 1997. *Liberal Peace, Liberal War: American Politics and International Security*. Ithaca, NY: Cornell University Press.
- Peceny, Mark, Caroline C. Beer, with Shannon Sanchez-Terry. 2002. "Dictatorial Peace?" *American Political Science Review* 96 (February): 15–26.
- Pitkin, Hanna Fenichel. 1972. *The Concept of Representation*. Berkeley: University of California Press.
- Przeworski, Adam, Michael Alvarez, Jose Antonio Chebub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990*. Cambridge: Cambridge University Press.
- Ray, James Lee. 1995. *Democracy and International Conflict: An Evaluation of the Democratic Peace Proposition*. Columbia: University of South Carolina Press.
- Ray, James Lee. 2003. "A Lakatosian View of the Democratic Peace Research Programme: Does It Falsify Realism (or Neorealism)?" In *Progress in International Relations Theory: Appraising the Field*, eds. Colin Elman and Miriam Fendius Elman. Cambridge, MA: MIT Press.
- Reiter, Dan, and Allan C. Stam. 2002. *Democracies at War*. Princeton, NJ: Princeton University Press.
- Reiter, Dan, and Allan C. Stam. 2003. "Understanding Victory. Why Political Institutions Matter." *International Security* 28 (Summer): 168–79.
- Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97 (November): 585–602.
- Rousseau, David, Christopher Gelpi, Dan Reiter, and Paul Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace." *American Political Science Review* 90 (September): 512–44.
- Rummel, R. J. 1995. "Democracies Are Less Warlike than Other Regimes." *European Journal of International Relations* 1 (December): 457–79.
- Russett, Bruce, and John Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: Norton.
- Russett, Bruce, and Harvey Starr. 1981. *World Politics: The Menu for Choice*. San Francisco: Freeman.
- Smith, Alastair. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43 (October): 1254–83.
- Vasquez, John A. 1997. "The Realist Paradigm and Degenerative versus Progressive Research Programs: An Appraisal of Neotraditional Research on Waltz's Balancing Proposition." *American Political Science Review* 91 (November): 899–912.
- Walt, Stephen M. 1987. *The Origins of Alliances*. Ithaca, NY: Cornell University Press.
- Waltz, Kenneth N. 1997. "Evaluating Theories." *American Political Science Review* 91 (November): 913–17.
- Zinnes, Dina A. 2004. "Constructing Political Logic: The Democratic Peace Puzzle." *Journal of Conflict Resolution* 48 (June): 430–54.

evaluation of the democratic peace project using several alternative metatheories, including Lakatos's (see also Ray 2003).

<sup>11</sup> Early indications from the quantitative literature are that hegemonic power does not have a pacifying effect—and, what is most important for democratic peace theory, the pacifying effect of dyadic democratic interaction is still in evidence after controlling for hegemonic power as well as other geopolitical factors highlighted by realist theory. Furthermore, this finding emerges from an analysis of the entire 20th century, not just the post-World War II period (Oneal and Russett 1999). Oddly, although Rosato (2003) cites them in support of his statement that democratic dyads were not any less likely to fight one another before the post-World War II period, Farber and Gowa (1997, 409) actually provide evidence that democracies were less likely to become embroiled in militarized disputes with each other as far back as 1919—and Gowa (1999, 99–100) acknowledges a similar finding for the decade preceding World War I.

# Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace

BRANISLAV L. SLANTCHEV *University of California—San Diego*

ANNA ALEXANDROVA *University of California—San Diego*

ERIK GARTZKE *Columbia University*

**R**osato (2003) claims to have discredited democratic peace theories. However, the methodological approach adopted by the study cannot reliably generate the conclusions espoused by the author. Rosato seems to misunderstand the probabilistic nature of most arguments about democratic peace and ignores issues that an appropriate research design should account for. Further, the study's use of case studies and data sets without attention to selection-bias produces examples that actually support theories it seeks to undermine. These problems place in doubt the article's findings.

**R**osato (2003) purports to demonstrate that the enormous literature on the democratic peace rests on dubious microfoundations. Reduced to its most basic, the claim is that none of the causal mechanisms advanced by the proponents of numerous different theories of the liberal peace hold up to empirical scrutiny. This is certainly an important finding if true. Unfortunately, the method employed in reaching these conclusions makes it impossible for us to know whether the author is right.

Despite the title of the article, the author does not engage the logic of the theories. Rather, he seeks to evaluate the empirical plausibility of the mechanisms they specify. We identify several problems with this methodology, each of which places in doubt the validity of the author's claims. Indeed, the study serves to catalogue research design flaws that are not uncommon in international relations research.

First, Rosato (2003) ignores fundamental issues of hypothesis testing and inference from historical data. We detail two possible interpretations of theoretical statements and show that the author's methodology does not allow him to draw the conclusions he does from either one. Second, the author ignores selection bias problems affecting observed behavior. This leads him to advance cases that actually support democratic peace theories instead of contradicting them.

We do not catalog all such errors, due to space constraints. Instead, we use the signaling theory (what Rosato refers to as "the information mechanism," 587) to illustrate most of our concerns.

## THE LOGIC OF INFERENCE: CAUSALITY AND EMPIRICAL TESTING

The most important errors in Rosato's article stem from inappropriate methodological choices and re-

search design. The basic setup of the study is a reduction of democratic peace theories to logical statements of implication of the form  $D \rightarrow S \rightarrow P$ , where  $D$  stands for "state is democratic,"  $S$  is a consequence implied by democracy (e.g., "state externalizes norms" or "state can signal better"), and  $P$  is the consequence of  $S$  (e.g., "states signaling or externalizing norms tend to resolve crises peacefully").<sup>1</sup>

Rosato (2003) seems to treat these statements as sufficient conditions. That is,  $D \rightarrow S$  means that democracy is all that is needed to achieve better signaling. The idea is to demonstrate that  $\neg[D \rightarrow S]$ , or that democracy does not imply the causal mechanism proposed by the theory. For example, Rosato (589) asserts that there are "several examples of liberal states violating liberal norms in their conduct of foreign policy and therefore the claim that liberal states generally externalize their internal norms of conflict resolution is open to question." In sentential logic, the argument boils down to  $\neg[D \rightarrow S] = [D \wedge \neg S]$ . Rosato reasons that if he demonstrates that  $[D \wedge \neg S]$  is true, then he can reject the claim that  $[D \rightarrow S]$ , which in turn negates the link between  $D$  and  $P$ . In other words, if he finds cases where a democracy ( $D$ ) failed to externalize norms ( $\neg S$ ), then he can infer that the causal connection postulated by the particular theory is empirically invalid and that the theory is thereby discredited.<sup>2</sup>

The problem with this reasoning is that democratic peace theories, as social scientific claims, do not typically offer hypotheses in the form of sufficient conditions. Instead, these theories make probabilistic claims for two reasons we explain in the following sections. We argue that Rosato's (2003) critique does not succeed irrespective of the source of the resulting empirical nondeterminism.

## EVALUATING THEORIES

Theoretical models express claims about *tendencies* that are contributions of one or several causal factors

Branislav L. Slantchev (slantchev@ucsd.edu) is Assistant Professor of Political Science, and Anna Alexandrova is a Ph.D. Candidate, Department of Philosophy, both at the University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0521 (aaalexan@ucsd.edu).

Erik Gartzke is Associate Professor of Political Science, Columbia University, New York, NY, 10027 (gartzke@columbia.edu).

We thank Hein Goemans, and Robert Northcott for useful discussions.

<sup>1</sup>  $D \rightarrow P$  means " $D$  implies  $P$ " (i.e.,  $D$  is a sufficient condition for  $P$  and  $P$  is a necessary condition for  $D$ ;  $\neg D$  means "not  $D$ "; and  $D \wedge S$  means " $D$  and  $S$ ").

<sup>2</sup> Rosato (2003) does not appear to challenge the  $S \rightarrow P$  component, at least in the cases we examine.

that would prevail and produce the anticipated effect *all other things being equal* (Hausman 1992, Mill 1967 [1836]). Take, for example, the signaling theory in Schultz (1998). The formal model demonstrates that public endorsement by the opposition tends to contribute positively (and, conversely, the absence of endorsement contributes negatively) to the credibility of the government's threat. The theory does *not* claim that (1) the opposition's actions will always (or even most of the time) lead to credible threats, or that (2) when a government's threats are credible, that this can be credited to the opposition. Liberal governments will make credible threats in the face of domestic dissent, even as they are bound to bluff occasionally, even when benefiting from domestic political consensus.

Because any theoretical model requires assumptions to produce its deductions, a careful theorist will be especially cautious in making predictions in cases where these assumptions may not hold; a judgment that is further complicated by the fact that we do not possess complete models and hence do not know the full set of assumptions that might be operating. The model expresses a tendency that should prevail in certain circumstances, but this tendency can also be overwhelmed by other, countervailing, ones. Anyone who seeks to assess a theory must make a reasoned judgment about where the theory applies. This requires that we identify a sample where the theory's assumptions are approximately satisfied. This would let the theory express a tendency claim about the real world rather than the neat stylized one of the model. Were one then to demonstrate that hypotheses from the theory do not obtain, one would have a serious challenge to the theory.

Rosato does not do this. Instead, he seeks to undermine democratic peace theory by selecting examples where the assumptions of theories are not satisfied, or where other factors held sway. For example, Rosato (2003, 589) challenges signaling theory in the following manner:

The available evidence suggests that democracies cannot clearly reveal their levels of resolve in a crisis. There are two reasons for this. First, democratic processes and institutions often reveal so much information that it is difficult for opposing states to interpret it.<sup>3</sup> Second, open domestic political competition does not ensure that states will reveal their private information.

The first sentence is demonstrably false. At least on occasion, democracies do appear to have been able to signal through open political contestation (see Schultz 1998). In addition, the two reasons Rosato gives for the alleged failure of democracies to signal are simply illustrations of countervailing tendencies. As such, Rosato's (2003) critique amounts to the rather unambi-

itious point that the theory applies in some cases more clearly than in others.

## PROBABILISTIC THEORIES

In drawing his conclusions, Rosato seems to treat theories as deterministic, whereas they are almost invariably couched in probabilistic terms. Theories in social science usually say things like "the probability of war is lower when informative signals can be sent" (Schultz 2001, 7), or "in any equilibrium of any game with the above format, the probability of war is an increasing function of the expected benefits from war of the informed player" (Banks 1990, 600).

Why couch theories in probabilistic terms? The probabilities in models can come from two sources. One of them is internal to models in the sense that a model may itself specify a probability distribution over outcomes arising from strategic factors. For example, it may be optimal to play a mixed strategy and bluff on occasion. Although we can specify the probability of bluffing, we cannot predict with certainty whether a player would bluff or not in any given realization of the game even if we hold everything else constant.

Another source of indeterminacy is external. Suppose the model itself makes a deterministic prediction. We still should not expect this prediction to hold once we "export" it to the empirical world. We simply cannot be sure how other factors, unforeseen by the theory, will play themselves out in individual cases. Because we do not have the complete specification of all contributing variables to social processes, we generally treat these unknowns as "noise." In testing, we seek to control for major disturbing factors (through case selection, multivariate statistical analysis, or experiment) and hope that the predicted tendency is robust enough to reveal itself regardless of other confounding influences.

Rosato (2003, 599) states that "the purported informational properties of democratic institutions are unlikely to improve the prospects for peace." The probabilistic claim that democracies do not lead to more credible signaling, and hence peace, is an assertion about statistical tendencies, not about behavior in individual cases, where outcomes can only occur or not occur. Though Rosato provides no carefully reasoned explication of the claim, let us assume that he is correct and that democracies do not strongly correlate with credible revelation of information. Suppose we found that out of five hundred interstate crises involving at least one democracy, only in 10% of the cases were democracies able to signal credibly, and in the remaining 90%, the tendency was supplanted by other causes. Is this democratic tendency then useless? The assertion that democracy does not explain anything would miss the point: after all, we may have a perfectly good explanation for 50 crises, and in the remaining cases, we may have a partial one. Focusing on the 90% of cases where the tendency was not decisive would mislead us to ignore the 10% where it was. Rejecting the theory on these grounds is unwarranted.

<sup>3</sup> The everyday use of the word "information" confuses the distinction between data (facts about defense spending, public statements, etc.) and private knowledge (e.g., one's reservation level). Rosato's (2003) claim appears to be that democracies make so much data available, that one would have difficulty inferring the privately known values from them. That is, he is saying that democracies do not reveal information, in the sense the concept is used in signaling games. We thank a reviewer for pointing this out.

Rosato's (2003) methodology, which fails for deterministic theories, is on even shakier ground for probabilistic claims. Under what conditions can we conclude that a tendency identified by a model is sufficiently causally relevant to explain outcomes in an appropriate sample of cases? Causality in these theories is not in the form of implications, but rather of probabilities. We say that  $D$  causes  $P$  if  $\Pr(P|S \wedge T) > \Pr(P|\neg S \wedge T)$  for every test situation  $T$ .<sup>4</sup> An appropriate test situation is one in which all other independent causally relevant factors are held fixed (Cartwright 1979). This condition was proposed to avoid Simpson's Paradox, where depending on how a population is partitioned a cause may actually decrease the probability of its effect.<sup>5</sup> We can interpret this as a requirement that the sample used for testing be chosen so as to respect the model's applicability. A researcher collects a sample of cases in which the model more or less applies and then measures the probability of its prediction coming true. Rosato's research design does not follow this widely accepted methodology for testing probabilistic hypotheses.

Because Rosato (2003) does not fully engage some of the theories he criticizes, the critique sometimes uses cases that actually support the theory he wants to discredit. Take, for example, the 1967 crisis between Egypt and Israel preceding the Six Days War. Citing Finel and Lord (1999), Rosato states that "Nasser was 'overwhelmed by the "noise" of Israeli domestic politics' and 'had enough information to see whatever he wanted and confirm existing misperceptions about Israeli intentions.'" This is said to illustrate how democracies cannot signal credibly.

Let us look at the tendencies the signaling theory expresses: democracies tend to signal credibly, and democratic signaling tends to decrease the probability of war. The hypothesis is that we are disproportionately unlikely to see democracies engaged in wars in cases where they are successful in signaling. Therefore, crises where for some reason the signaling tendency is overwhelmed by other factors are more likely to end in war. The theory leads us to expect that crises that involve democracies and that end in war are precisely the ones where democracies failed to reveal information through signaling. Rosato's (2003) example refers to just such a crisis and thus lends support to the theory.

<sup>4</sup>  $\Pr(P|S \wedge T)$  reads "probability of event  $P$  conditional on events  $S$  and  $T$  occurring jointly."

<sup>5</sup> Suppose that democracies signal more credibly but also tend to be weak militarily. If credible signaling is a cause of peace, but military weakness is an even greater cause of war (by inviting attack), then democracies may appear more likely to end up at war than nondemocracies. If  $S$  represents credible signaling and  $M$  represents military weakness,  $\Pr(P|S) > \Pr(P|\neg S)$ . However, if we condition on whether the military is weak, the inequality is reversed:  $\Pr(P|S \wedge M) < \Pr(P|\neg S \wedge M)$  and  $\Pr(P|S \wedge \neg M) > \Pr(P|\neg S \wedge \neg M)$ . These reversals constitute Simpson's Paradox (Hitchcock 2002). The requirement that only independent causal factors are held fixed is also necessary. Suppose that some cause  $M$  of  $P$  is itself caused by  $S$ . If  $S$  causes  $P$  exclusively through  $M$ , then holding  $M$  fixed would screen off  $S$  from  $P$ , something we clearly want to avoid.

## SELECTION BIAS

One must be careful in using cases presumably produced by the data-generating process that the models are trying to explain. Selection bias in conflict datasets has been a well-known problem for some time, and researchers are typically at pains to ensure that they account for its misleading effects. In particular, one must infer the consequences of a theory for observable behavior or else risk reaching incorrect conclusions.

Take, for example, the theory that democratic leaders are more readily punished if they lose a war, and hence that they are more reluctant to engage in wars, making democracies less likely to escalate crises to the highest level of violence. Rosato (2003, 594) uses Goemans (2000) data of the fates of leaders after war "to determine whether leaders' decisions for war are affected by their domestic accountability, that is, if there is something about the domestic structure of states that affects their chances of being punished."

According to the theory, leaders take into account the chances of being punished if they lose, and the fear of punishment affects their conflict decisions. Therefore, cases where war actually occurs already tend to contain leaders who have discounted the probability of punishment. Suppose that democratic leaders who lose a war are more likely to be punished than autocratic ones (we are not saying that this is true; we are just conducting a thought experiment). It follows that democratic leaders would tend to get involved only in wars they believe they can win; hence, democracies would tend to win the wars they fight (this is what we observe empirically). What happens in the few cases where democratic leaders lose? As Rosato (2003) himself finds, these leaders tend to get removed from office disproportionately.

Rosato (2003, 594) concludes that "this evidence is not strong. This is because there are only four cases of democratic losers in the entire dataset, making it impossible to draw any firm conclusions about the likelihood that losing democrats will be removed." But this conclusion is clearly wrong, for, according to the logic of the argument, the evidence is overwhelmingly in support of the self-selection hypothesis: few democracies lose, and in those cases that democracies do lose, leaders get removed at very high rates. We would conclude that (1) democratic leaders are, in fact, more likely to be removed if they lose, and therefore (2) they would only fight when the chances of losing are sufficiently small, and so (3) we should observe very few cases where democratic leaders lose wars. Similar arguments apply to costly wars: after all, few leaders would deliberately begin wars that they expect to be costly and long.

## CONCLUSION

The method Rosato (2003) uses to discredit democratic peace theories is inappropriate in most social science contexts. Because Rosato's article is a manifestation of a widespread misconception in our discipline, we

believe it is worth drawing attention to the problems inherent in such approaches.

Despite the title of his article, Rosato does not engage the logic of the theories he wants to discredit. We are willing to believe that many explanations for the democratic peace offer internally inconsistent or *ad hoc* arguments. For many of these theories, it is an open question under what assumptions their claims hold. However, using historical examples to challenge logic is misleading; we know neither that the logic of the theory is correct nor that the implications of the theory are wrong. We suspect, for example, that any reasonably competent student of history can interpret a given case in various ways to support contradictory hypotheses.

Without a proper evaluation of the logic of competing theories, one might (charitably) assume equal deductive consistency for all. We would then hope to see a demonstration that some theories are less useful empirically than others. Instead, Rosato (2003) offers yet another theory: American preponderance, principally through NATO, is said to explain the democratic peace. But this theory needs a proper empirical evaluation missing from the article.<sup>6</sup>

We believe that progress in social science is best achieved through an interactive simultaneous advance on two fronts: the construction of internally consistent theories and the careful comparative empirical evaluation of competing models. If Rosato's (2003) critique of democratic peace theory fails to strike its target, it stands to do substantial damage by legitimizing a fundamentally incorrect method of evaluating social science theories. Although scholars with normative

aversion to the democratic peace or the scientific method may conclude that their views have been vindicated, we hope to have demonstrated that such a conclusion cannot depend on Rosato's study.

## REFERENCES

- Banks, Jeffrey S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34 (August): 599–614.
- Cartwright, Nancy. 1979. "Causal Laws and Effective Strategies" *Nous* 13: 419–37.
- Finel, Bernard L., and Kristin M. Lord. 1999. "The Surprising Logic of Transparency." *International Studies Quarterly* 43 (June): 315–39.
- Gartzke, Erik. 2004. "The Futility of War. Capitalism and Common Interests as Determinants of the Democratic Peace." Manuscript, Department of Political Science, Columbia University.
- Goemans, Hein E. 2000. "Fighting for Survival: The Fate of Leaders and the Duration of War." *The Journal of Conflict Resolution* 44 (October): 555–79.
- Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hitchcock, Christopher. 2002. Probabilistic Causation. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. URL=<http://plato.stanford.edu/entries/causation-probabilistic/>.
- Mill, John Stuart. 1967 [1836]. "On the Definition of Political Economy and the Method of Investigation Proper to It." In *Collected Works of John Stuart Mill*. Vol. 4. Toronto: University of Toronto Press.
- Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97 (November): 585–602.
- Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92 (December): 829–44.
- Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.

<sup>6</sup> Rosato's (2003) hypothesis is not supported by a large-N analysis: Adding joint NATO membership in a dyad as a dummy variable to standard statistical models of the democratic peace does not alter the effects of democracy, and is itself statistically insignificant (Gartzke 2004). The hypothesis is easily refuted even by Rosato's own approach to testing: The peace observation holds for non-NATO dyads (Austria-Switzerland) and fails for NATO partners (Greece-Turkey).

# Three Pillars of the Liberal Peace

MICHAEL W. DOYLE *Columbia University*

**S**ebastian Rosato (2003) finds the logic of the “democratic peace” flawed in his “*The Flawed Logic of Democratic Peace Theory*,” and he cites my work and other studies as examples of the flawed logic. Some of the logic he describes is flawed, and it may characterize some of the literature in the wide field of “democratic peace,” but it is not the logic underlying the core of liberal peace theory. Indeed, the persuasive core of the logic underlying the theory of liberal democratic peace is missing from Rosato’s account. Republican representation, an ideological commitment to fundamental human rights, and transnational interdependence are the three pillars of the explanation. The logic underlying the peace among liberal states rests on a simple and straightforward proposition that connects those three causal mechanisms as they operate together and only together, and not separately as Sebastian Rosato claims.

**I** explain the persuasive core of the logic underlying the theory of liberal democratic peace logic in three places. The two-part essay “Kant, Liberal Legacies and Foreign Affairs” published in *Philosophy and Public Affairs* (1983) showed how Immanuel Kant’s (1790) 1795 essay, “Perpetual Peace,” could be constructed as a coherent explanation of two important regularities in world politics—the tendencies of liberal states simultaneously to be peace-prone in their relations with each other and war-prone in their relations with nonliberal states. Republican representation, an ideological commitment to fundamental human rights, and transnational interdependence are the three causal mechanisms of the explanation. These are Kant’s three “definitive articles”—the constitutional, international and cosmopolitan laws—of the hypothetical peace treaty he asks states to sign. The first part of the two-part essay focuses on the liberal peace and its Kantian sources. The second part of the two-part essay focuses on exposing the dangers of liberal imperialism, liberal aggression and liberal appeasement (Rosato 1996 cites the reprints of the two articles in *Debating the Democratic Peace*). I also addressed these themes in the *American Political Science Review* in December 1986 and distinguished Kantian “liberal internationalism” from “liberal pacifism” and “liberal imperialism.” In 1997, in *Ways of War and Peace*, I distinguished liberalism from the two other major traditions of international thought, Realism and Marxism.

All three have one consistent and key argument: “No one of these constitutional, international or cosmopolitan sources is alone sufficient, but together (and only where together) they plausibly connect the characteristics of liberal politics and economies with sustained liberal peace” ([1983a, 1983b] 1996, 27). I repeat the same sentence as the summary of the argument—“No single constitutional, international . . .” in the *American Political Science Review* (Doyle 1986, 1162); and the identical sentence (this time in italics for emphasis) in *Ways of War and Peace* (1997, 284). I explicitly

say that the three causes explained liberal peace and liberal war when, and only when, combined. Rosato’s critique of the work, nonetheless, rests on treating each of these factors—“normative,” “institutional,” etc.—in isolation as if they were sufficient.

This is important because, in my view, no one of the factors alone is a sufficient explanation of the liberal peace or liberal war. First, as Rosato correctly suggests, there is no reason for all direct or indirectly majoritarian governments to be peaceful toward other majoritarian governments. Clearly, a democracy of xenophobes or hyper-nationalists would externalize their preferences. Anticipating Rosato’s critique, I (1997, chaps. 4 and 9) pointed out in *Ways of War and Peace*—in chapters that discuss Rousseau and Marx—that democratic institutions are completely compatible with Realist foreign policy when preferences are integrally and exclusively nationalist and with Socialist solidarity and international class warfare when strictly egalitarian (and societies lack individual liberties and private property). Jean Jacques Rousseau’s classic account of democratic theory, for example, anticipates that democracies will be locked, as any Realist would agree, in a generalized “state of war” with all other states, whether democratic or not (Rousseau 1756/1917). If information flows across borders are limited, subject to manipulation and nationalist myth-making, and each democracy cultivates a normative commitment to complete autonomy and self-help, democracies will be likely to clash (Mearsheimer 1990; VanEvera 1990).

Second, there should be no expectation that a population widely sharing liberal values associated with human rights norm will shape policy unless they have democratic representation with the transparency and accountability that can shape public decision-making.<sup>1</sup>

And third, there is no guarantee that commercial and other forms of interdependence will alone provide material foundations for cooperation among societies, rather than for sources of imperial rivalry and fuel to balance of power competition, unless trade and investment are part of a relationship of trust and respect.<sup>2</sup>

Michael W. Doyle is Harold Brown Professor, School of International and Public Affairs and Columbia Law School, Columbia University, 1314 IAB, 420 West 118th Street, New York, NY 10027. (md2221@columbia.edu).

The author thanks Robert Jervis and Frank Wayman for advice on an earlier draft of this note

<sup>1</sup> Mueller (1989) stresses the norms of peace in his explanation, but also links these norms to democratic institutions.

<sup>2</sup> Cobden (1901) is a classic source on the pacifying effects of trade. Russett and O’Neal (2001) and Gartzke, Li, and Boehmer (2001)



Rosato (2003) is right to criticize the logic of each of these strands of the "democratic peace" standing alone.

Once combined, however, the three sources do help explain why liberal states maintain peace with each other and are, nonetheless, and for explicable reasons prone to war and imperialism with nonliberal states. Given the absence of this explanation in Mr. Rosato's (2003) critique, it is worth briefly summarizing the three hypotheses here. Each emphasizes one aspect of what characterizes a liberal republic.

First, republican representative democratic governments tend to create an accountable relationship between the state and the voters, particularly median voters. They preclude monarchs or dictators turning their potentially aggressive interests into public policy while assuming that the costs will be borne by a subordinate public. Democratic representation introduces republican caution, Kant's (1970) "hesitation," in place of autocratic caprice. Representative government allows for a rotation of elites. This encourages a reversal of disastrous policies as electorates punish the party in power with electoral defeat. Legislatures and public opinion further restrain executives from policies that clearly violate the obvious and fundamental interests of the public, as the public perceives those interests.

As importantly, representation together with transparency (what Kant [1970] called "publicity") may provide for effective signaling, assuring foreign decision makers that democratic commitments are credible because rash acts and exposed bluffs will lead to electoral defeat. Able to make more credible commitments than regimes with more narrow selectorates, democracies would thus be less likely to stumble into wars.<sup>3</sup>

We should not, however, overemphasize rational signaling. The division of powers and rotation of elites characteristic of republican regimes can permit mixed signals, allowing foreign powers to suspect that executive policies might be overturned by legislatures, courts, or the next election. On the other hand, the shared powers of republics should encourage better chances for deliberation. Most importantly, the combination of representative institutions and purely rational material interests do not control for the possibility that powerful states can have rational incentives to conquer wealthy and exploit wealthy, weak democracies. If reputations are short and differentiable and supposedly pacifying long-run interests are indeterminate, as they often are, something more than rational material interest will be needed to explain liberal peace.

Representation should, however, ensure that liberal wars are only fought for popular, liberal purposes. This does not produce peace. The historical liberal legacy is laden with popular wars fought to promote freedom, protect private property or support liberal allies against nonliberal enemies.<sup>4</sup> In order to see how the

pacific union removes the occasion of wars among liberal states and not wars between liberal and nonliberal states, we need to shift our attention from liberal representation to liberal principles and liberal interests, the other two elements in the liberal explanation of peace and war. These latter two elements account for the purposes that representative processes promote and what credible signaling needs to signal.

Second, liberal principles add the prospect of international respect. Liberal principles, or norms, involve an appreciation of the legitimate rights of all individuals. Connecting these principles to public policy requires publicity. Domestically, publicity helps ensure that the officials of republics act according to the principles they profess to be just and according to the interests of the electors they claim to represent. Internationally, free speech and the effective communication of accurate conceptions of the political life of foreign peoples are essential to establish and preserve the understanding on which the guarantee of respect depends.

These principles begin the differentiation of policy toward liberal and nonliberal states, requiring trust of and accommodation toward fellow liberals and producing distrust of and opposition toward nonliberals. Domestically just republics, which rest on the consent of free individuals, presume foreign republics to be also consensual, just, and therefore deserving of the accommodation that the individuals that compose them deserve. The experience of cooperation helps engender further cooperative behavior when the consequences of state policy are unclear but (potentially) mutually beneficial. At the same time, liberal states assume that nonliberal states, which do not rest on free consent, are not just. Because nonliberal governments are perceived to be in a state of aggression with their own people, their foreign relations become, for liberal governments, deeply suspect. In short, fellow liberals benefit from a presumption of amity; nonliberals suffer from a presumption of enmity. Both presumptions may be accurate. Each, however, may also be self-fulfilling.

Democratic liberals do not need to assume either that public opinion rules foreign policy or that the entire governmental elite is liberal. They can assume that the elite typically manages public affairs but that potentially nonliberal members of the elite have reason to doubt that antiliberal policies would be electorally sustained and endorsed by the majority of the democratic public.

Third and last, material incentives sustain interliberal normative commitments. The "spirit of commerce" spreads widely and creates incentives for states to promote peace and to try to avert war. Liberal economic theory holds that these cosmopolitan ties derive from a cooperative international division of labor and free

draw links between trade and peace, but only in the context of wider relationships favoring accommodation

<sup>3</sup> For further discussion, see Fearon 1994, Gaubatz 1996; Schultz 1998, and Lipson 2003.

<sup>4</sup> This is the theme of Small and Singer (1976) and Doyle (1983b), Chan (1984), and Weede (1984). Many liberal philosophers, includ-

ing Kant in "Perpetual Peace," regard these wars as unjust, and Kant warns liberals of their susceptibility to them (see 1970, 106). At the same time, he argues that each nation "can and ought to" demand that its neighboring nations enter into the pacific union of liberal states (102) whose first requirement is domestically liberal institutions.

trade according to comparative advantage when the parties can expect to be governed by a rule of law that respects property and that enforces legitimate exchanges. Each economy is said to be better off than it would have been under autarky; each thus acquires an incentive to avoid policies that would lead the other to break these economic ties. But, because keeping open markets rests on an assumption that the next set of transactions will also be determined by prices rather than coercion, a sense of mutual security is vital to avoid security-motivated searches for economic autarky. Thus, avoiding a challenge to another liberal state's security or even enhancing each other's security by means of alliance naturally follows economic interdependence.

In this same regard, a further cosmopolitan source of liberal peace is that the international market removes difficult decisions of production and distribution from the direct sphere of state policy. A foreign state thus does not appear directly responsible for these outcomes; states can stand aside from, and to some degree above, these contentious market rivalries and be ready to step in to resolve crises. The interdependence of commerce and the international contacts of state officials help create cross-cutting transnational ties that serve as lobbies for mutual accommodation. According to modern liberal scholars, international financiers and transnational and transgovernmental organizations create interests in favor of accommodation. Moreover, their variety has ensured no single conflict sours an entire relationship by setting off a spiral of reciprocated retaliation.

Conversely, the suspicion that characterizes relations between liberal and nonliberal governments can lead to restrictions on the range of contacts between societies. And this can increase the prospect that a single conflict will determine an entire relationship. As importantly, in relations with weak societies, "protecting 'native rights' from native oppressors, and protecting universal rights of property and settlement from local transgressions, introduced especially liberal motives for imperial rule" (Doyle [1983a, 1983b] 1996, p. 37). When property lacks clear title and exchanges are subject to manipulation and uncertain legal enforcement—the typical environment of non-liberal states—then economic contact generates strife.

No single constitutional, international, or cosmopolitan source alone is sufficient. This variant of liberal theory is neither solely institutional, nor solely ideological, nor solely economic. But together (and only together) the three specific strands of liberal institutions, liberal ideas, and transnational ties plausibly connect the characteristics of liberal polities and economies with sustained liberal peace. But in their relations with nonliberal states, liberal states have not escaped from the insecurity caused by anarchy in the world political system considered as a whole. Moreover, the very constitutional restraint, international respect for individual rights, and shared commercial interests that establish grounds for peace among liberal states establish grounds for additional conflict in relations between liberal and nonliberal societies.

Thus, when Rosato (2003, 588, 593) criticizes the "norm externalization" argument or the "institutional logic" explanations, he is in each case missing two thirds of the liberal argument. When he argues that economic interests and local strategic interests shaped liberal imperial policy in the 19th century, he is not refuting—he is confirming—the logic of the liberal peace. In these cases, principled liberal motives joined material interests in liberal imperialism. Campaigns against the slave trade destabilized commercial oligarchies, making them prone to collapse. The *mission civilatrice* and the "dual mandate" imperial ideologies both included liberal principles, albeit ones that allowed for liberal imperial paternalism of the sort J. S. Mill (1859/1973) endorsed for societies he and his fellow liberals saw as incapable of governing themselves. But commercial and property interests, which lacked institutionalization in much of Africa and Asia, were even more important. Lacking both legal recognition and the context of interliberal respect, commercial and property claims fueled imperialism. (Doyle [1983a, 1983b] 1996, 37–9). Liberals were all too ready to enforce those property claims both as a matter of material interest and principled defense of rights. Interliberal peace rests on the combined effect of the three pillars. Absent one of them, pacific policy is underdetermined and undermined.

During the Cold War, the United States did intervene against or take measures to undermine covertly numerous popular regimes in the Third World. In many cases the U.S. administration in office was convinced that the regimes in question (Mossadegh in Iran, Arbenz in Guatemala, Jagan in Guyana, Allende in Chile, and the Sandinistas in Nicaragua) were threats both to property and to the rule of law. The fact that these regimes were more progressive and popular than any previous regime in those countries (and, in some cases, since) did not make them well-established liberal democracies. Many U.S. officials doubted their stability as democracies. They were also seen as influenced by and allied with communist regimes. President Kennedy articulated the logic clearly, referring to the assassination of Trujillo in the Dominican Republic: "There are three possibilities in descending order of preference, a decent democratic regime, a continuation of the Trujillo regime or a Castro regime. We ought to aim at the first, but we cannot really renounce the second until we are sure that we can avoid the third" (Schlesinger 1965, 769, quoted in Doyle [1983] 1996, 41). As importantly, all of these interventions were covert; they lacked the mechanisms of publicity on which the liberal peace rests. The explanation underlying the liberal peace makes no assumption that every official, always and everywhere, is motivated by liberal principle and interest—just that over the normal political cycle nonliberal principles and interests will not become the norm in the formation of liberal foreign policy.

A much more logical explanation comes with methodological costs. Data sets on the liberal peace do not adequately code for these three pillars together and separately. My own coding (1983a, 1983b, 1986, 1997)

was approximate. The most thorough recent empirical test of Kantian propositions (Russett and O'Neal 2001) shows the separate positive effects of democratic institutions and trade (and membership in international organizations), but it doesn't separately code for liberal norms. The substantial statistical confirmation that inter-democratic peace, (coding for democratic institutions), does receive is thus probably a reflection of the tendency for principles of liberal individualism and democratic institutions to evolve together.<sup>5</sup> But we cannot be sure of this. Compared to other testable international theories of similar scope, the empirical confirmation of the liberal peace is exceptionally strong, but that does not mean that the theory does not need additional testing.

## REFERENCES

- Chan, Steve. 1984. "Mirror, Mirror on the Wall . . . Are the Freer Countries More Pacific." *Journal of Conflict Resolution* 28:4 (December): 617-48.
- Cobden, Richard. 1901. *Political Writings of Richard Cobden*. London: Ridgway.
- Doyle, Michael. 1983a. "Kant, Liberal Legacies, and Foreign Affairs. Part I." *Philosophy and Public Affairs* 12, 3 (June): 205-35.
- Doyle, Michael. 1983b. "Kant, Liberal Legacies, and Foreign Affairs: Part II." *Philosophy and Public Affairs* 12, 4 (October): 323-53.
- Doyle, Michael. 1986. "Liberalism and World Politics." *American Political Science Review* 80: 4 (December): 1151-69.
- Doyle, Michael. 1996. "Kant, Liberal Legacies, and Foreign Affairs." In *Debating the Democratic Peace*, ed. Michael Brown, Sean Lynn-Jones, and Steven E. Miller. Cambridge: MIT Press, 3-57.
- Doyle, Michael. 1997. *Ways of War and Peace*. New York: Norton.
- Fearon, James. 1994. "Domestic Political Audiences and the Escalation of Political Disputes." *American Political Science Review* 88, 3 (September): 577-92.
- Gartzke, Eric, Quan Li, and Charles Boehmer. 2001. "Investing in the Peace." *International Organization* 55 (Spring), 391-438.
- Gaubatz, Kurt. 1996. "Democratic States and Commitment in International Politics." *International Organization* 50.1 (Winter): 109-39.
- Kant, Immanuel. 1790. "Perpetual Peace. In *Kant's Political Writings*, ed. Hans Reiss, trans. H. B. Nisbet. Cambridge, UK: Cambridge University Press.
- Lai, Brian and Dan Reiter. 2000. "Democracy, Political Similarity and International Alliances." *Journal of Conflict Resolution* 44 (April).
- Lipson, Charles. 2003. *Reliable Partners*. Princeton: Princeton University Press.
- Mearsheimer, John. 1990. "Back to the Future." *International Security* 15, 1 (Summer): 5-56.
- Mill, John Stuart. 1859/1973. "A Few Words on Nonintervention." In *Essays on Politics and Culture*, ed. Gertrude Himmelfarb. Gloucester, MA: Peter Smith.
- Mueller, John. 1989. *Retreat from Doomsday*. Basic Books.
- Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97: 4 (November): 585-602.
- Rousseau, Jean-Jacques. 1756/1917. "State of war" in C. E. Vaughan, ed. *A Lasting Peace through the Federation of Europe*. London: Constable, 1917.
- Rummel, R. J. 1997. *Power Kills*. New Brunswick, NJ: Transaction Publishers.
- Russett, Bruce, and John O'Neal. 2001. *Triangulating Peace*. New York: Norton.
- Schlesinger, Arthur. 1965. *A Thousand Days*. Boston: Houghton Mifflin.
- Schultz, Kenneth. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92: 4 (December): 829-844.
- Small, Melvin, and J. David Singer. 1976. "The War Proneness of Democratic Regimes." *Jerusalem Journal of International Relations* 1, 4 (Summer): 50-69.
- Van Evera, Stephen. 1990. "Primed for Peace." *International Security* 15, 1: 58-107.
- Weede, Eric. 1984. "Democracy and War Involvement." *Journal of Conflict Resolution* 28.4 (December): 649-64.

<sup>5</sup> For thorough surveys of empirical literature see Rummel (1997) and Lai and Reiter (2000).

# Explaining the Democratic Peace

SEBASTIAN ROSATO *The University of Chicago*

*I am grateful for the opportunity to respond to the rejoinders to my article, "The Flawed Logic of Democratic Peace Theory" (Rosato 2003). In each case, I summarize the core issues at stake and explain why I do not believe that my critics have succeeded in casting serious doubt on my original argument.*

## KINSELLA ON CAUSAL LOGICS AND THEORY REJECTION

### Monadic Logics

David Kinsella's (2005) first major claim is that my criticism of democratic peace theory is misdirected because I test the theory's causal logics as if they are monadic when they are in fact dyadic. Evidence from conflicts between democracies and non-democracies is irrelevant, he argues, because the logics state that democracies will only externalize their domestic norms of conflict resolution or act cautiously in conflicts with other democracies.

What Kinsella fails to realize is that although the democratic peace *finding* is dyadic, the *logics* adduced to explain it are monadic. The six logics that I identified in my article all begin with the claim that democratic norms and institutions cause democracies to behave differently from nondemocracies in systematic ways: there are fewer reasons available to them for going to war, they are more constrained in the use of violence, they are slower to resort to force, and they are better at signaling their levels of resolve. In essence, the argument is that democracies are less violence-prone than are other kinds of states and/or more effective at engaging in the kind of behavior that makes war less likely. Proponents of the democratic peace then use these monadic tendencies to explain why democracies have not fought one another. Simply put, in a crisis involving two democracies, each side has a low propensity for violence and a high aptitude for the kind of behavior that makes war less likely, and each knows that its democratic opponent also has these qualities. Therefore, they are able to remain at peace (Bueno de Mesquita et al. 1999; Russett 1993; Schultz 2001).<sup>1</sup>

Sebastian Rosato is a Ph.D. Candidate, Department of Political Science, The University of Chicago, 5828 South University Avenue, Chicago, IL 60637. He is also a Fellow, John M. Olin Institute for Strategic Studies, 1033 Massachusetts Avenue, Cambridge, MA 02138. (srosato@uchicago.edu)

I thank Alexander Downes, John Mearsheimer, John Schuessler, Robert Trager, and my colleagues at the Olin Institute for their comments and suggestions.

<sup>1</sup> The perceptual versions of these logics are also monadic. They take the following form. Democracy A is constrained and, because it perceives State B as a democracy, believes B is also constrained. B carries out the same calculation. Thus, A and B are able to remain at peace. Doyle's (1997) claim that democracies remain at peace because they trust and respect one another and fight nondemocracies because they neither trust nor respect them is the only example of democratic peace theorists proposing a dyadic logic. As I explained in my article, however, this logic is *ad hoc* (Rosato 2003, 589–90).

Let me approach this same point from a slightly different perspective. The logics underpinning the democratic peace refer to how democracies act with respect to all states, whether democratic or not. The public constraint logic, for example, states that pairs of democracies remain at peace because both parties face above average constraints in deciding to go to war with any adversary, not just with other democracies. Similarly, the information logic suggests that members of democratic dyads do not fight because they are both good at signaling their level of resolve, not because they are only good at signaling other democracies. In short, democratic peace theory's logics rest on a "multiplier" argument: if a state with a low propensity for violence comes into contact with another state that also has a low propensity for violence, then the likelihood of war breaking out is very low indeed.<sup>2</sup>

Therefore, in order to evaluate democratic peace theory's logics, we must determine whether democratic norms and institutions actually cause democracies to behave differently from nondemocracies in systematic ways. For example, is there good evidence that democracy causes greater elite accountability, better access to the policy process for peace-loving interest groups, better signaling in crises, and a greater commitment to the use of peaceful norms of conflict resolution? If there is, then we have a plausible explanation for the democratic peace finding. If not, then the peace that exists among democracies may not be caused by the democratic nature of those states.

This is the kind of evaluation that I carried out in my article before concluding that democracy does not have the effects that proponents of the democratic peace attribute to it (Rosato 2003, 599). Liberal democracies do not reliably externalize their domestic norms of conflict resolution. Democratic leaders are not especially accountable to peace-loving publics or pacific interest groups. Democracies are not particularly slow to mobilize or incapable of surprise attack. And open political competition offers no guarantee that a democracy will reveal private information about its level of resolve. Therefore, the existing logics cannot explain the democratic peace finding: two democracies, each relatively unconstrained and expecting the other to be similarly unconstrained, may well fight one another.

In sum, the logics underpinning democratic peace theory are monadic in form; thus the tests that I carried

<sup>2</sup> Similarly, if a state that can effectively reveal private information about its level of resolve comes into contact with another state that can do the same, then the likelihood that they will fight is quite low.

out provide good evidence that the absence of war between democracies may not be caused by their democratic nature.

### Theory Rejection

Kinsella's other major claim is that the criteria that I adopted to reject democratic peace theory are unfair. He argues, first of all, that I cannot reject the theory on the basis of a handful of select historical examples that belie its causal logics. Moreover, he faults me for claiming that democratic peace theory is a degenerating research paradigm because scholars in that tradition focus on perceptions. I agree that had I adopted either of these approaches, then my critique of democratic peace theory would have been inadequate. However, I used neither strategy in my article.

**Selected Cases.** Rather than relying on a few examples to show that democratic peace theory's causal logics *occasionally* fail to play out as advertised, I used large numbers of cases to show that the causal mechanisms *often* fail to operate as stipulated. Moreover, I tested the logics on sets of cases that were most likely to support democratic peace theory. My reasoning was that if there was little evidence that the logics operated in these "easy" cases, then this would cast serious doubt on the theory. That said, Kinsella is right to note that I did not make either point explicit in my article. A brief summary and evaluation of my findings is therefore in order.<sup>3</sup>

In examining the argument that democracies generally externalize their domestic norms of conflict resolution, I identified 33 wars in which they failed to do so. In each case, I looked for evidence that the war in question could plausibly be justified on the grounds of self-defense or the inculcation of liberal values and found that it could not. I also argued that there may be up to 33 more wars in which democracies attempted to perpetuate or reimpose autocratic rule in direct violation of their domestic norms of conflict resolution. In the case of the trust and respect logic, I cited a total of 18 examples of democracies failing to trust and respect one another. Because every case involved a pair of democracies, they should have lent support to the logic rather than contradicting it. My analysis of the group constraint mechanism found that prowar groups in the United States and Britain have often prevailed over antiwar groups in domestic debates during the last two centuries. Similarly, an analysis of U.S. foreign policy decisions since 1789 suggests that American presidents have been able to circumvent or overcome checks and balances almost at will. My decision to focus on Britain and the United States when evaluating these mechanisms was intentional: if the logics fail to operate in the most democratic of states, then they are likely to fare even worse in states that are less democratic. Finally, in the case of the public constraint mechanism, I showed that the logic failed to operate as stipulated in 12 of

15 cases where democratic peace theorists would most expect it to apply. In addition, I cited a dozen crucial examples where democratic publics appear to have imposed no constraints on their leaders even though the other state was democratic (Rosato 2003, 588–99).

These findings would cast doubt on any set of causal logics, but they are especially damaging to democratic peace theory because its principal finding holds that democracies have rarely if ever fought one another. If democratic peace theory's causal logics are to explain this finding, then they should rarely fail to operate as stipulated. But they appear to fail fairly frequently, and we therefore have reason to doubt their explanatory power.

**Perceptions.** I did not argue that democratic peace theorists' attempts to repair their logics by introducing perceptions are an indication that the research program is degenerating. If there is good evidence that, in order to remain at peace, states must not only be democratic but also perceive one another as such, then a focus on perceptions is entirely appropriate. In other words, I agree with Kinsella that the turn to perceptions need not be an indication that the research program is degenerating. In fact, Kinsella appears to acknowledge this, noting that I did not use the term "degenerating research program" in my evaluation of democratic peace theory's causal logics.

My point about perceptions was different. In essence, I argued that bringing in perceptions can only improve a logic's power if we can predict how democracies will categorize other states with a high level of confidence and if this categorization is relatively stable. I then provided evidence that strategic interest or policymakers' personal beliefs and party affiliations have often prevented democracies from forming coherent, accurate, and stable assessments of other states' regime type, thereby lessening our confidence that joint democracy can enable democracies to remain at peace. Moreover, I argued that democratic peace theorists have failed to come up with a compelling theory of perceptions; they cannot tell us when democracies will perceive other states as democratic and when they will not (Rosato 2003, 592–93). Because Kinsella disputes neither my reasoning nor my findings, I find his critique unconvincing.

### SLANTCHEV, ALEXANDROVA, AND GARTZKE ON SCIENTIFIC INQUIRY

#### Probabilistic Causality and the Information Logic

Branislav Slantchev, Anna Alexandrova, and Erik Gartzke's (2005) first criticism of my article is that I mistakenly treat theories as if they are deterministic rather than probabilistic and that I evaluate them on that basis.

I agree that social science theories are probabilistic: they are designed to simplify reality and, in the course of simplifying, theorists are bound to sacrifice some explanatory power. It is for this reason that I

<sup>3</sup> I deal with the accountability and information mechanisms in my response to Slantchev, Alexandrova, and Gartzke below.

chose to cast doubt on the causal logics by citing large numbers of anomalies rather than selected historical cases. I reasoned that this would allow me to claim that the logics *rarely* operated as stipulated or were *frequently* overwhelmed by other factors. Although such an approach cannot decisively disconfirm probabilistic logics, it can suggest that their explanatory power is highly circumscribed.

My analysis of the information logic explicitly recognizes the fact that it is probabilistic and demonstrates that it *frequently* fails to operate as advertised.<sup>4</sup> The logic states that opposition-party support tends to contribute positively to the credibility of a democracy's threat, whereas lack of support contributes negatively to the credibility of a threat. In response, I argued that opposition party support *rarely* contributes positively to the credibility of a threat because it is what we *expect* opposition parties to do. There are several reasons why support for the government is likely to be the default strategy, including "rally round the flag" effects, nationalism, and elite control over relevant information. Schultz's (2001) data provide evidence for this claim: democratic governments that have issued deterrent threats have received opposition-party support 84% of the time. In short, the fact that a democracy's opposition party supports the government rarely conveys information during a crisis because this is what the other state expects it to do (Rosato 2003, 598–99).

The important fact to note about *opposed* threats is that they are rare. This should not surprise us because, as I have just noted, opposition parties will overwhelmingly support their governments. This means that we need only cite a handful of examples where opposition parties opposed the use of force but governments went to war anyway in order to cast doubt on the logic. This is what I did in my article (Rosato 2003, 599). Alternatively, we can identify crises in which an opposition party opposed a deterrent threat—as Schultz does—and check to see whether deterrence failed more often than it succeeded. Contrary to what democratic peace theorists would expect, we find that the opposite is true: deterrence succeeded in three of the five cases (Schultz 2001, 167). In sum, there are good reasons to believe that democracies are not especially good at conveying information about their levels of resolve. Most of the time they convey little if any information, and on the rare occasions that they do convey information, that information does not appear to exert a substantial impact on crisis outcomes.

Although I identified several cases where the information logic does not apply or does not operate as stipulated, Slantchev, Alexandrova, and Gartzke argue that it cannot be rejected because there are still some cases that it *can* explain. What fraction of a given set of cases must a logic explain for us to accept it? My critics are prepared to endorse a logic with a 10% success rate.

I do not find this argument convincing for two reasons. First, although I agree that theories are probabilistic, it is not clear to me that we should be satisfied with a logic that has a 10% success rate. This is not to say that it is useless, but we should make every effort to come up with logics that explain a larger fraction of the empirical record. A related issue here is the question of falsifiability. All logics can explain at least some cases because scholars generate theories from their observation of historical events (Powell 1999). If we note this fact and couple it with the claim that even theories that explain a small percentage of cases are useful, we are in effect arguing that theories cannot be falsified: all theories can explain a few cases (the cases that they are based on) and theories that can explain a few cases cannot, apparently, be thrown out.

Second, a theory with a 10% success rate is hardly satisfying if we consider that democracies have rarely fought one another. Instead, we are left wondering what other factors are at work in bringing about this result. A possible fallback position here would be the claim that there are several logics associated with democracy, and although each logic only explains a fraction of the cases, they explain most of the cases when taken together. The implication of this argument would be that democracy is a "master variable" that explains the democratic peace through several causal mechanisms. We should, however, be wary of claims such as this one. Any research program can presumably proliferate logics that explain a fraction of the cases from a single master variable, but were we to adopt this approach we would simply be engaging in "curve-fitting" exercises rather than coming up with powerful logics that propose simple explanations for large numbers of cases.

There is, however, no need to engage in a debate about the requirements of a good theory to make my point. Recall that my central claim about the information logic is that democracies are not especially good at revealing their levels of resolve in a crisis because the stance taken by opposition parties rarely sends an informative signal. This implies that if we conduct a statistical test of the kind recommended by Slantchev, Alexandrova, and Gartzke, then we should find little support for the information logic. In order to evaluate this proposition, I took the cases of attempted deterrence that Schultz used to test the information logic in his own work and carried out a probit analysis to determine whether the stance taken by opposition parties correlates with the probability of deterrence success. I included one control variable—the balance of power—based on my intuition that states are more likely to deter potential attackers if they are more powerful than they are and less likely to do so if they are weaker.

According to the results, neither "supported democratic defender" nor "opposed democratic defender" are significant at the 5% level (Table 1). The coefficient on "balance of power" is, however, both large and significant ( $p = 0.01$ ). These results suggest that (1) opposition party support or lack of support is not significantly associated with the probability of deterrence

<sup>4</sup> I focus on Schultz's (2001) information logic in order to reply directly to Slantchev, Alexandrova, and Gartzke. Schultz himself argues that his contribution to democratic peace theory is suggestive rather than conclusive.

**TABLE 1. Probability of Deterrence Success (Probit Estimates)**

Variable	Coefficient	Standard Error
Constant	-0.85	0.40
Supported democratic defender	0.69	0.40
Opposed democratic defender	0.24	0.62
Balance of power	1.08*	0.44
$\chi^2$	10.63*	
N	57	

Notes. \* $p < 0.05$ . I thank Kenneth Schultz for providing me with his data. I coded balance of power using Slinger and Small 1993. In order to determine whether the attacker or defender was more powerful, I first added their total military personnel and calculated the percentage of that total accounted for by the attacker and defender. I did the same for military expenditure, steel production, and electricity consumption. Then I averaged together each state's percentages for personnel, expenditure, steel, and electricity and determined which of the two possessed a greater share of their total power. Like Schultz (2001) I calculated Huber-White robust standard errors and clustered cases within the same crisis. The dataset is available upon request.

success, and (2) threats made by democratic governments and supported by opposition parties are no more likely to succeed than are threats by nondemocracies.<sup>5</sup> In short, as I argued in my article, democracy does not appear to be associated with better signaling.

### Selection Bias and the Accountability Logic

Slantchev, Alexandrova, and Gartzke's other major criticism is that my analysis of the accountability logic is plagued by selection bias, which leads me to cite evidence that supports the logic rather than discrediting it. I am puzzled by the accusation of selection bias. I did not select cases on the dependent variable in my analysis, and my critics give no evidence that I did so.

This methodological quibble aside, the evidence in my article casts significant doubt on the accountability logic. According to Slantchev, Alexandrova, and Gartzke, my finding that democratic leaders are more likely than autocratic leaders to be removed from office for losing a war lends credence to democratic peace theorists' claims that democrats are more accountable than are autocrats. In my article, however, I argued that accountability is determined not only by the probability of removal, but also by the costs that leaders will incur in the event they are removed from office. These costs include imprisonment, exile and death or, simply, "punishment." Thus I argued (Rosato

2003, 593–94) that leaders make decisions based on *expected costs*. Slantchev, Alexandrova, and Gartzke do not dispute this claim; they simply ignore it.

What do we see when we factor in costs? Using a well-known dataset I compared the fates of democratic and autocratic leaders who took their countries into costly or losing wars. In the case of costly wars, there is little debate: autocrats are both more likely to be removed and more likely to be punished. Losing wars provide a more complicated picture. On the face of it, democratic losers are removed 75% of the time, whereas autocratic leaders are removed 35% of the time. But as I argued in my article, we should not count the Menzies resignation as an example of removal and therefore democrats are more likely to be removed 50% to 35%. Autocrats are, however, far more likely to be punished (29% to 0%). Because democrats are more likely to be removed and autocrats are more likely to be punished, I argued that we cannot claim that either are more accountable (Rosato 2003, 594).

There is now more evidence for my claims. Chiozza and Goemans (2004) use a dataset of all leaders between 1919 and 1999 to determine whether defeat in war affects the tenure of democratic and nondemocratic leaders. Their findings are stronger even than mine: defeat in war significantly reduces the tenure of nondemocratic leaders, but does not significantly affect the tenure of democratic leaders. In other words, autocrats know that war involvement can reduce their time in power, and democrats know that war involvement has little if any effect on their chances of retaining power. In sum, the evidence does not support the claim that democrats are more accountable than autocrats.

Faced with these findings, my critics shift their position on the accountability issue. Their new argument goes as follows. If we assume that democrats are more likely than autocrats to be punished for losing a war, then it follows that democrats will only get into wars that they can win and will therefore win most of the wars that they fight. Slantchev, Alexandrova, and Gartzke then note that democracies do indeed win most of their wars and assert that this must be because democratic leaders are more accountable than their autocratic counterparts.

This argument is unconvincing. The problem is that, as I have shown, there is scant evidence for the initial premise of my critics' new argument. Slantchev, Alexandrova, and Gartzke are wrong to assume that democrats are more likely to be punished for losing a war and are therefore more accountable than autocrats. Therefore, they cannot assert that democrats will only get into wars that they can win and will consequently win most of the wars that they fight. This is not to say that democracies do not win a lot of wars—there is good evidence that they do—but their war-winning cannot be attributed to their greater accountability.

### Evaluating Theories

Slantchev, Alexandrova, and Gartzke do more than simply question the persuasiveness of my critiques:

<sup>5</sup> I ran another probit that included the 10 independent variables that Schultz (2001) used with one exception: I replaced his balance of forces variable with my balance of power variable. Neither supported nor opposed democratic defender were significant at the 5% level, whereas the coefficient on the balance of power variable was both large and significant ( $p < 0.001$ ). We must treat these results with caution, because they rest on a sample of only 57 cases. At a minimum, however, we can conclude that support for the information logic is not robust.



my article, they argue, legitimizes “a fundamentally incorrect method of evaluating social science theories.” For them a causal logic ought to be evaluated in two ways. First, we must establish whether it is logically consistent. Second, we must determine whether, all else equal, its independent and dependent variables are correlated. Take their restatement of the accountability logic: they argue that it must be considered powerful because it is logically consistent and because democracies win most of the wars that they fight (democracy and war-winning are highly correlated).

My approach to theory testing is different. In addition to checking for logical consistency and correlation, I seek to establish whether the logic actually operates as stipulated (Rosato 2003, 585–86). Where is the evidence that democratic leaders think and act in accordance with the logic and choose easy wars for fear of losing office if they are defeated? In short, where is the evidence that the relationship is *causal* rather than merely *correlational*?

Despite Slantchev, Alexandrova, and Gartzke's claims, this debate cannot be won by asserting that their testing method is scientific while mine is not—both of our approaches have a scientific basis (MacDonald 2003). Indeed, democratic peace theorists appear to be gravitating toward my way of doing business. Having established that there is a correlation between joint democracy and peace, they have turned to the task of developing a set of causal logics connecting the two variables.<sup>6</sup> If they are successful and we find good evidence that these logics actually operate as stipulated, then their theory must be considered compelling. However, as I argued in my article, the logics that they have provided so far do not work as advertised; therefore, the democratic peace continues to be an empirical finding in search of an explanation.

## DOYLE'S THREE PILLARS

According to Michael Doyle (2005), my article ignores his seminal claim that democracies remain at peace because they are simultaneously cautious, respectful towards one another, and committed to promoting peace among themselves. I do not doubt that states that are fundamentally cautious, respect each other, and want to remain at peace, will remain at peace. Instead, my claim was that democratic norms and institutions do not reliably cause caution and respect, and therefore cannot be the cause of the peace that exists among democracies.

Doyle's explanation for the democratic or liberal peace rests on three logics. The first logic states that democratic institutions and processes “create an accountable relationship between the state and the voters.” This in turn induces “caution” in the international arena because there are a variety of circumstances in which voters—broadly defined to include the general public, interest groups and legislatures—are likely to

oppose war. According to the second logic, elites in democracies “act according to the principles they profess to be just,” assume that other democracies are also just, and therefore respect one another. The institutional and normative logics that I describe in my article are identical to these two logics (Rosato 2003, 586–87). Doyle's third logic holds that a basic commitment to liberal economic norms encourages a “spirit of commerce” among democracies, which in turn impels them to promote peace and try to avert war with one another. I did not lay out or test a logic analogous to this one since most democratic peace theorists focus on regime type and ignore economic interdependence. Moreover, as I demonstrate below, the fact that I ignored this “third pillar” does not weaken my claims.

Having elaborated these three logics, Doyle goes on to argue that they operate “together and only together” to bring about peace between democracies. In other words, the democratic peace finding exists because all three of the following obtain simultaneously: democracies are fundamentally cautious about using force, democracies respect one another, and democracies work hard to promote peaceful relations with fellow democratic states. It follows—and Doyle is explicit about this—that if any one of these factors does not obtain, then we should not expect to see peace among democracies.

I did not question this core argument in my article. In fact, I agree that two states that are fundamentally hesitant to use force, respect one another, and work to remain at peace will rarely if ever fight one another. I am also satisfied with the claim that in the absence of one of these factors states may well fight one another.

My argument was different: I checked the historical record to see whether there is good evidence that democratic institutions do indeed induce caution and whether a domestic commitment to democratic norms does indeed cause states to respect one another. In other words, I did not ask whether caution plus respect causes peace; rather, I asked whether democracy reliably causes caution and respect. I found that it does not. Democratic leaders do not appear to be especially accountable to peace-loving publics or pacific interest groups, therefore casting doubt on the claim that democracy induces caution (Rosato 2003, 593–99). Similarly, there is substantial evidence that democracies do not reliably externalize their domestic norms of conflict resolution and do not respect one another when their interests clash (Rosato 2003, 588–93). In sum, democracy does not reliably induce caution or respect and, crucially, rarely causes both simultaneously.

By Doyle's own reasoning this finding means that we should see several wars between democracies and, because democracies appear to act little differently from nondemocracies, as many wars between democracies as between other kinds of states. Yet democracies have rarely if ever fought one another and have created a separate peace. There is, in short, a mismatch between the outcome predicted by Doyle's logic and what we actually observe in the world. The source of this mismatch is obvious: having discovered that democracies

<sup>6</sup> Note, however, that the finding itself has recently come under attack (Henderson 2002).

are consistently peaceful in their relations with one another, Doyle has explained the finding with a set of criteria that, as I demonstrated in my article, do not reliably obtain separately and rarely obtain simultaneously.

Curiously, Doyle's critique actually undermines democratic peace theory by making it harder to validate. When I wrote my article, I reasoned that I would have to show that neither the institutional logic nor the normative logic operated as stipulated. Doyle's rejoinder, however, makes it clear that in order to cast doubt on democratic peace theory we need only find evidence that one logic rarely operates as advertised. If democracy does not reliably lead to caution, for example, then he would predict at least a handful of wars between democracies. Because democracies have not fought one another a handful of times, his argument falls short.

## CONCLUSION

My purpose in writing "Flawed Logic" was to cast doubt on the logics underpinning the democratic peace. I do not find the criticisms leveled at the piece convincing and stand by my claim that, although there is peace among democracies, it does not appear to be caused by the democratic nature of those states. Nevertheless, I did not intend or expect to have the last word on the subject. Rather, my intention from the start was to spark a debate about the most important liberal theory of war and peace. I thank my critics for joining that debate and hope that others will follow suit.

## REFERENCES

- Bueno de Mesquita, Bruce, et al. 1999. "An Institutional Explanation of the Democratic Peace." *American Political Science Review* 93 (December): 791-812.
- Chiozza, Giacomo, and H. E. Goemans. 2004. "International Conflict and the Tenure of Leaders: Is War Still *Ex Post* Inefficient?" *American Journal of Political Science* 48 (July): 604-19.
- Doyle, Michael W. 1997. "Kant, Liberal Legacies, and Foreign Affairs." In *Debating the Democratic Peace*, ed. Michael E. Brown, Sean M. Lynn-Jones, and Steven E. Miller. Cambridge, MA: MIT Press.
- Doyle, Michael W. 2005. "Three Pillars of the Liberal Peace." *American Political Science Review* 99 (August): 463-66.
- Henderson, Errol A. 2002. *Democracy and War: The End of an Illusion?* Boulder, CO: Lynne Rienner.
- Kinsella, David. 2005. "No Rest for the Democratic Peace." *American Political Science Review* 99 (August): 453-57.
- MacDonald, Paul K. 2003. "Useful Fiction or Miracle Maker: The Competing Epistemological Foundations of Rational Choice Theory." *American Political Science Review* 97 (November): 551-66.
- Powell, Robert. 1999. "The Modeling Enterprise and Security Studies." *International Security* 24 (Fall): 97-106.
- Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97 (November): 585-602.
- Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post Cold War World*. Princeton, NJ: Princeton University Press.
- Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.
- Singer, J. David, and Melvin Small. 1994. Correlates of War Project: International and Civil War Data, 1816-1992 [computer file] (Study 9905). Ann Arbor, MI: J. David Singer and Melvin Small [producers], 1993. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1994.
- Slantchev, Branislav L., Anna Alexandrova, and Erik Gartzke. 2005. "Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace." *American Political Science Review* 99 (August): 459-62.



*American Political Science Association*

## **American Political Science Review**

*Volume 99 | Number 4 | November 2005*

### **INSIDE THIS ISSUE:**

#### NOTES FROM THE EDITOR

#### ARTICLES

- Majority Norms, Multiculturalism, and Gender Equality  
**Sarah Song** ..... 473
- To Trust an Adversary: Integrating Rational and Psychological Models in Collaborative Policymaking  
**William D. Leach and Paul A. Sabatier** ..... 491
- Using Geographic Information Systems to Study Interstate Competition  
**William D. Berry and Brady Baybeck** ..... 505
- Ethnicity and Trust: Evidence from Russia  
**Donna Bahry, Mikhail Kosolapov, Polina Kozyreva, and Rick K. Wilson** ..... 521
- Military Coercion in Interstate Crises  
**Branislav L. Slantchev** ..... 533
- Contracting around International Uncertainty  
**Barbara Koremenos** ..... 549
- Centripetal Democratic Governance: A Theory and Global Inquiry  
**John Gerring, Strom C. Thacker, and Carola Moreno** ..... 567
- "Interpose Your Friendly Hand": Political Supports for the Exercise of Judicial Review by the United States Supreme Court  
**Keith E. Whittington** ..... 583
- Political Science on the Cusp: Recovering a Discipline's Past  
**John G. Gunnell** ..... 597

#### FORUM

- Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance  
**Jana von Stein** ..... 611
- The Constraining Power of International Treaties: Theory and Methods  
**Beth A. Simmons and Daniel J. Hopkins** ..... 623

APSR EXTERNAL REVIEWERS, 2004–2005 ..... 633

INDEX TO VOLUME 99 ..... 639



## EDITOR

Lee Sigelman  
*Department of Political Science  
The George Washington University  
Washington, DC 20052*

## EXECUTIVE COMMITTEE

### American Politics

Robert Stein

*Rice University*

### Comparative Politics

Thomas Remington

*Emory University*

### International Relations

Edward Mansfield

*University of Pennsylvania*

### Political Theory

Ruth W. Grant

*Duke University*

### At-Large

K C Morrison

*University of Missouri*

M. Elizabeth Sanders

*Cornell University*

## EDITORIAL BOARD

Martha Ackelsberg

*Smith College*

James Adams

*University of California,  
Santa Barbara*

Kathleen Bawn

*University of California,  
Los Angeles*

Janet Box-Steffensmeier

*The Ohio State University*

Cathy Cohen

*University of Chicago*

Timothy E. Cook

*Louisiana State University*

Neta Crawford

*Brown University*

Ivor Crewe

*University of Essex*

Darren Davis

*Michigan State University*

Jon Elster

*Columbia University*

Diana Evans

*Trinity College*

Steven Finkel

*University of Virginia*

Kim L. Fridkin

*Arizona State University*

Robert Goodin

*Australian National  
University*

Peter A. Hall

*Harvard University*

Richard Herrmann

*The Ohio State University*

Rodney Hero

*University of Notre Dame*

James Johnson

*University of Rochester*

Jane Junn

*Rutgers University*

George Klosko

*University of Virginia*

Arthur Lupia

*University of Michigan*

Forrest Maltzman

*The George Washington  
University*

Kenneth Mayer

*University of Wisconsin*

Kirstie McClure

*University of California,  
Los Angeles*

Kenneth Meier

*Texas A&M University  
and Cardiff University*

James D. Morrow

*University of Michigan*

Rebecca Morton

*New York University*

Robert Price

*University of California,  
Berkeley*

Kathryn Sikkink

*University of Minnesota*

Beth Simmons

*Harvard University*

Jack Snyder

*Columbia University*

Janice Gross Stein

*University of Toronto*

Sven Steinmo

*University of Colorado*

Shannon Stimson

*University of California,  
Berkeley*

Leslie Paul Thiele

*University of Florida*

Daniel Treisman

*University of California,  
Los Angeles*

Clyde Wilcox

*Georgetown University*

John Wright

*The Ohio State University*

Bernard Yack

*Brandeis University*

Langche Zeng

*The George Washington  
University*

## EDITORIAL STAFF

Elizabeth Cook

*Assistant Editor*

Jennifer Deets

Elizabeth Franker

Lee Michael

Jennie Schulze

*Editorial Assistants*

**Association Office Address and Membership Information:** American Political Science Association individual membership dues are Regular Members with income \$100,000+, \$190, \$80,000–\$99,999, \$163, \$60,000–\$79,999, \$146, \$50,000–\$59,999, \$136, \$40,000–\$49,999, \$115, Under \$40,000, \$78, Student Members \$37; Retired over \$25,000, \$52, Retired under \$25,000, \$32; Life Members, \$3,000 00, Unemployed members, \$37 00, Associate Member with one journal selection, \$48. The amount of the dues allocated for a subscription is \$5 for an individual membership. Changes of address for members should be completed online or mailed to the APSA membership office at APSA, 1527 New Hampshire Avenue, NW, Washington, DC 20036.

**Subscription, Publishing, and Advertising Office Address:** Cambridge University Press, 40 West 20th Street, New York, NY 10011-4211; and (outside the United States, Canada, and Mexico) Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU, England.

**Subscription Information:** *American Political Science Review* (ISSN 0003-0554) is published quarterly in February, May, August, and November by Cambridge University Press for the American Political Science Association. Annual institutional electronic-only subscription rate (2005) is US\$425 in the United States, Canada, and Mexico, UK£268 elsewhere. Annual institutional print-and-electronic subscription rate (2005) is US\$520 in the United States, Canada, and Mexico, UK£322 elsewhere. Periodicals postage rate paid at New York, NY and additional mailing offices. Institutional subscription orders and address changes should be sent to *American Political Science Review*, Cambridge University Press, 100 Brook Hill Drive, West Nyack, NY 10994-2133, USA.

Copyright © 2005 by the American Political Science Association

All rights reserved. No part of this publication may be reproduced, in any form or by any means, electronic, photocopying, or otherwise, without permission in writing from Cambridge University Press. Photocopying information for users in the U.S.A.: The Item-Fee Code for the publication (0003-0554/05 \$9.00 + \$10) indicates that copying for internal or personal use beyond that permitted by Sec. 107 or 108 of the U.S. Copyright Law is authorized for users duly registered with the Copyright Clearance Center (CCC) Transaction Reporting Service, provided that the appropriate remittance of \$9.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. Specific written permission must be obtained for all other copying.

# American Political Science Review

Volume 99, Number 4, November 2005

## Table of Contents

NOTES FROM THE EDITOR .....	iii
ARTICLES	
Majority Norms, Multiculturalism, and Gender Equality <b>Sarah Song</b> .....	473
To Trust an Adversary: Integrating Rational and Psychological Models in Collaborative Policymaking <b>William D. Leach and Paul A. Sabatier</b> .....	491
Using Geographic Information Systems to Study Interstate Competition <b>William D. Berry and Brady Baybeck</b> .....	505
Ethnicity and Trust: Evidence from Russia <b>Donna Bahry, Mikhail Kosolapov, Polina Kozyreva, and Rick K. Wilson</b> .....	521
Military Coercion in Interstate Crises <b>Branislav L. Slantchev</b> .....	533
Contracting around International Uncertainty <b>Barbara Koremenos</b> .....	549
Centripetal Democratic Governance: A Theory and Global Inquiry <b>John Gerring, Strom C. Thacker, and Carola Moreno</b> .....	567
"Interpose Your Friendly Hand": Political Supports for the Exercise of Judicial Review by the United States Supreme Court <b>Keith E. Whittington</b> .....	583
Political Science on the Cusp: Recovering a Discipline's Past <b>John G. Gunnell</b> .....	597
FORUM	
Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance <b>Jana von Stein</b> .....	611
The Constraining Power of International Treaties: Theory and Methods <b>Beth A. Simmons and Daniel J. Hopkins</b> .....	623
APSR EXTERNAL REVIEWERS, 2004-2005 .....	633
INDEX TO VOLUME 99 .....	639

## Officers

**PRESIDENT**  
**Margaret Levi**  
University of Washington

**PRESIDENT-ELECT**  
**Ira Katznelson**  
Columbia University

**VICE-PRESIDENTS**  
**Robert Axelrod**  
University of Michigan

**John A. Garcia**  
University of Arizona

**Joan C. Tronto**  
CUNY Graduate Center and  
Hunter College

**SECRETARY**  
**Christine Marie Sierra**  
University of New Mexico

**TREASURER**  
**Henry E. Brady**  
University of California,  
Berkeley

**PROGRAM CO-CHAIRS**  
**James D. Johnson**  
University of Rochester

**Jack Knight**  
Washington University,  
St. Louis

**Susan C. Stokes**  
University of Chicago

**EDITOR, APSR**  
**Lee Sigelman**  
George Washington University

**EDITOR, PERSPECTIVES**  
**James Johnson**  
University of Rochester

**EXECUTIVE DIRECTOR**  
**Michael A. Brintnall**  
American Political Science  
Association

## Council

### 2003–2005

Tony Affigne  
Providence College

John H. Aldrich  
Duke University

Manuel Avalos  
Arizona State University West

Judith A. Baer  
Texas A&M University

Shirley M. Geiger  
Savannah State University

John W. Harbeson  
CUNY-City College of  
New York

Marion E. Orr  
Brown University

Joanna Vecchiarelli Scott  
Eastern Michigan University

### 2004–2006

Lisa Anderson  
Columbia University

Andrew L. Aoki  
Augsburg College

Neta C. Crawford  
Brown University

James Louis Gibson  
Washington University,  
St. Louis

Harvey C. Mansfield  
Harvard University

Rogers M. Smith  
University of Pennsylvania

Aili Tripp  
University of Wisconsin,  
Madison

David Vogel  
University of California,  
Berkeley

## Former Presidents

Frank J. Goodnow  
Albert Shaw  
Frederick N. Judson  
James Bryce  
A. Lawrence Lowell  
Woodrow Wilson  
Simeon E. Baldwin  
Albert Bushnell Hart  
W. W. Willoughby  
John Bassett Moore  
Ernst Freund  
Jesse Macy  
Munroe Smith  
Henry Jones Ford  
Paul S. Reinsch  
Leo S. Rowe  
William A. Dunning  
Harry A. Garfield  
James W. Garner  
Charles E. Merriam  
Charles A. Beard  
William Bennett Munro  
Jesse S. Reeves  
John A. Fairlie  
Benjamin F. Shambaugh

Edward S. Corwin  
William F. Willoughby  
Isidor Loeb  
Walter J. Shepard  
Francis W. Coker  
Arthur N. Holcombe  
Thomas Reed Powell  
Clarence A. Dykstra  
Charles Grove Haines  
Robert C. Brooks  
Frédéric A. Ogg  
William Anderson  
Robert E. Cushman  
Leonard D. White  
John Gaus  
Walter F. Dodd  
Arthur W. MacMahon  
Henry R. Spencer  
Quincy Wright  
James K. Pollock  
Peter H. Odegard  
Luther Gulick  
Pendleton Herring  
Ralph J. Bunche  
Charles McKinley

Harold D. Lasswell  
E. E. Schattschneider  
V. O. Key Jr.  
R. Taylor Cole  
Carl B. Swisher  
Emmette S. Redford  
Charles S. Hyneman  
Carl J. Friedrich  
C. Herman Pritchett  
David B. Truman  
Gabriel A. Almond  
Robert A. Dahl  
Merle Fainsod  
David Easton  
Karl W. Deutsch  
Robert E. Lane  
Heinz Eulau  
Robert E. Ward  
Avery Leiserson  
Austin Ranney  
James MacGregor Burns  
Samuel H. Beer  
John C. Wahlke  
Leon D. Epstein  
Warren E. Miller

Charles E. Lindblom  
Seymour Martin Lipset  
William H. Riker  
Philip E. Converse  
Richard F. Fenno Jr.  
Aaron B. Wildavsky  
Samuel P. Huntington  
Kenneth N. Waltz  
Lucian W. Pye  
Judith N. Shklar  
Théodore J. Lowi  
James Q. Wilson  
Lucius J. Barker  
Charles O. Jones  
Sidney Verba  
Arend Lijphart  
Elinor Ostrom  
M. Kent Jennings  
Matthew Holden Jr.  
Robert O. Keohane  
Robert Jervis  
Robert D. Putnam  
Theda Skocpol  
Susanne Hoerber Rudolph

## Notes from the Editor

Each year at this time it is my pleasure to acknowledge the contributions that hundreds of reviewers have made to the *APSR* and, through it, to the profession. The individuals whose names are listed in "APSR External Reviewers 2004–2005" later in this issue served as reviewers—some of several papers—between mid-August 2004 and mid-August 2005. They have my sincere gratitude for their service, sine qua non.

### IN THIS ISSUE<sup>1</sup>

Notwithstanding "great man" theories of history, effective political action—a successful revolution, an orderly implementation of a new policy, and so on—generally requires efforts that extend well beyond those of an isolated individual. This issue's cover image of a bridge visually fixes the common thread of interconnectedness that runs through the first four of our otherwise wide-ranging set of November articles. In these articles, cultures collide in the courtroom, advocates argue environmental policy, states vie for a competitive edge, and neighboring cultures learn to coexist—but never alone. Be it in societies, interest groups, or political jurisdictions, interests and preferences have far-reaching effects, reshaping the distribution of political winners and losers, reallocating resources and bragging rights, and redefining friends and foes.

Is it wrong to protect or accommodate racial or ethnic minority groups when doing so can imperil the rights of women within those minorities? Although many have posed the issue as one of multiculturalism versus gender equality, Sarah Song doubts that the matter is that clear-cut. In "Majority Norms, Multiculturalism, and Gender Equality," Song recommends scrutiny of minority groups' cultural claims, consideration of the biases of the majority culture, and monitoring for harmful spillover effects that accommodation might create. Analyzing controversies involving Indians' tribal membership, immigrants' criminal defenses, women's citizenship rights, and Mormons' polygamy, Song shows that American history has not been confined to instances in which the majority culture has condemned minority cultural practices, but also has offered examples of how each side can support, encourage, adopt, or overshadow biases in the practices of the other. More broadly, Song's thought-provoking article highlights how cultures change, for better or for worse, over time and in response to their surroundings.

Whereas cultures often just bump into each other, interest groups are often statutorily required to interact in certain policy arenas. In "To Trust an Adversary: Integrating Rational and Psychological Models of Collaborative Policymaking," William D. Leach and Paul A. Sabatier explore two different perspectives—a

rational choice-based approach and a psychological one—to explain the factors that enabled members of watershed stakeholder partnerships in the American West to trust one another and work together on controversial environmental policies. Whereas the rational choice approach suggests a tit-for-tat model of trust-building based on the availability of information and of monitoring institutions capable of applying sanctions, psychological models focus on participants' beliefs, cognitive limitations, and perceptions of the legitimacy of the process. Rather than pitting the two explanatory approaches against one other, Leach and Sabatier allow for the possibility that they may operate jointly. The payoff comes when a welter of interview and survey data indicates that each model conveys insights into how these policy elites were able to build trust and work together. Leach and Sabatier's findings not only provide an example of how knowledge can be built on multiple theoretical bases, but help us understand real-life situations in which unlikely allies find themselves able to cooperate.

Interconnectedness is about more than winning culture wars or policy contests. Considerations of learning and economic advantage stand out in "Using Geographic Information Systems to Study Interstate Competition." William D. Berry and Brady Baybeck use geographic information systems, a new set of tools for political scientists, to some old questions: Do states learn from each other? Do they compete with each other? Berry and Baybeck reassess two often-studied state-level policy issues, lottery adoption and the generosity of welfare benefits, via this new technique, which treats states as geographic spaces with nodes of varying population densities. Just as different-sized planets and stars exert varying amounts of "pull" on other objects in space, new techniques allow for the possibility that states like California and Montana exert differing levels of influence on their neighbors. These and related GIS technologies should be useful in studying not only interactions among American states, but also subnational politics elsewhere and policy diffusion at the international level.

Donna Bahry, Mikhail Koslapov, Polina Kozyreva, and Rick K. Wilson tear down the proposition that "good fences make good neighbors," in "Ethnicity and Trust: Evidence from Russia." Based on data from surveys in Tartarstan and Sakha-Yakutia, Bahry and her colleagues conclude that the amount of interaction among different ethnic groups and trust in government are the strongest indicators of inter-group trust. This novel finding has important implications for questions of group identity and interpersonal trust in multi-ethnic societies, particularly regarding the link between in-group trust and out-group trust, which the authors conclude are not inversely related. Their counterintuitive conclusion that generalized trust is not the best predictor of inter-group trust should be of considerable interest to a wide range of scholars who focus on

<sup>1</sup> Drafted by editorial assistant Lee Michael.



issues of interconnectedness, such as collective action, ethnicity, and nationalism.

The balance of this issue ranges far and wide, from levying war and concluding peace through domestic politics and institutional design to a bit of political science history as well. The topical smorgasbord begins with another round of war and peace scholarship.

Albert Einstein believed that "You cannot prevent and prepare for war at the same time." In "Military Coercion in Interstate Crises," Branislav L. Slantchev challenges this notion by showing how military mobilization can deter an opponent in a crisis situation. Rather than relying on classic arguments about audience costs, Slantchev uses an elegant model to demonstrate how military mobilization simultaneously ties the hands of politicians and sinks costs, doubly signaling the mobilizer's resolve. This innovative treatment of tacit bargaining during crises directly challenges the contention of democratic peace theory advocates that democracies are better able to signal their intentions because they face higher audience costs. Slantchev's contention that autocracies are able to signal their intentions as well as democracies when military means are available to them is likely to spark several additional rounds of debate on the causes of war and peace.

When the fighting stops, peace is inaugurated with paperwork: treaties and other international agreements are often considered to be long-lasting guarantees of behavior and obligations between signatories. Visions of parchment, quill pens, and elaborate signing ceremonies in gilded halls or on carrier decks that usher in new eras of cooperation come to mind. However, Barbara Koremenos' research on international agreements on economics, the environment, human rights, and security, as reported in "Contracting around International Uncertainty," reveals that states more often than not make multiple short-term arrangements in the face of an uncertain international environment. Koremenos' analysis should be of interest not only to the international relations scholars, but also to others with interests in institutions and institutional design, including both Americanists and comparativists.

Those Americanists and comparativists will already be interested in identifying constitutional structures that give rise to "good government." John Gerring, Strom C. Thacker, and Carola Moreno take a broad view of this question in "A Centripetal Theory of Democratic Governance: A Theory and Global Inquiry," based on debates about presidentialism versus parliamentarism, federalism versus unitarism, and single-member districts versus proportional representation. Gerring and associates believe that the latter types of institutions, which form the basis of centripetalism, facilitate higher standards of living and good governance compared to states with vertical and horizontal separations of power. In this sense, that government governs best which governs most—an argument that promises to reignite the debate about whether and in what ways centralized authority and broad inclusiveness are superior means to democratic ends.

Another important question about democratic ends concerns the role of the courts in democratic decision making. While judicial review is often interpreted as an assault on the policy making prerogatives of elected officials, Keith E. Whittington's "Interpose your Friendly Hand": Political Supports for the Exercise of Judicial Review by the United States Supreme Court" explores how courts serve the political and electoral needs of the dominant national coalition in overcoming barriers to implementing their political agenda. Using episodes of judicial review by the U.S. Supreme Court as case studies, Whittington sets out to determine when elected officials might find it advantageous to pursue policy and electoral objectives through the judiciary. The result is a novel contribution that should be read not only by public law scholars, but by Americanists and comparativists who too often ignore the policy making role of the courts.

Asked when political science shifted toward its modern-day embrace of "science," most political scientists would probably identify the turning point as the "behavioral revolution" of the mid-twentieth century. However, John G. Gunnell, in "Political Science on the Cusp: Recovering a Discipline's Past," argues that the changes of the 1950s and '60s were more like an academic reformation than a discipline-altering revolution. The turning point, according to Gunnell, took place during the 1920s. The true founding fathers of modern political science were scholars like G. E. G. Caitlin and W. Y. Elliott, whose works initiated a paradigm shift in political science. Gunnell provides evidence that these then-prominent but now largely forgotten figures deserve a more prominent place in our discipline's annals than they have received to date. (Our publication of this article serves the secondary function of providing another occasion to make known that our November 2006 issue will complete the one-hundredth volume of the *APSR*. As previously announced, our centennial issue will be given over to articles on the theme of the evolution of political science.)

In the December 2000 *APSR*, Beth A. Simmons argued in "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs" that reputational concerns lead states to comply with their treaty obligations. In the "Forum" section of this issue, Jana von Stein contends in "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance" that selection bias problems mask states' true motivation for obeying treaty obligations. The key factor, von Stein argues, is the set of conditions that led them to sign the treaty in the first place, not their concern about how other states would respond if they were to shirk. In "The Constraining Power of International Treaties: Theory and Methods," Simmons and Daniel J. Hopkins question the robustness of von Stein's findings, recast Simmons' model to mitigate von Stein's methodological concerns, and conclude that Simmons' original results still hold. This exchange ends here so far as the *APSR* is concerned, but research on the vital question of treaties and state behavior will surely continue.

## INSTRUCTIONS TO CONTRIBUTORS

### General Considerations

The *APSR* strives to publish scholarly research of exceptional merit, focusing on important issues and demonstrating the highest standards of excellence in conceptualization, exposition, methodology, and craftsmanship. Because the *APSR* reaches a diverse audience of scholars and practitioners, authors must demonstrate how their analysis illuminates a significant research problem, or answers an important research question, of general interest in political science. For the same reason, authors must strive for a presentation that will be understandable to as many scholars as possible, consistent with the nature of their material.

The *APSR* publishes original work. Therefore, authors should not submit articles containing tables, figures, or substantial amounts of text that have already been published or are forthcoming in other places, or that have been included in other manuscripts submitted for review to book publishers or periodicals (including on-line journals). In many such cases, subsequent publication of this material would violate the copyright of the other publisher. The *APSR* also does not consider papers that are currently under review by other journals or duplicate or overlap with parts of larger manuscripts that have been submitted to other publishers (including publishers of both books and periodicals). Submission of manuscripts substantially similar to those submitted or published elsewhere, or as part of a book or other larger work, is also strongly discouraged. If you have any questions about whether these policies apply in your particular case, you should discuss any such publications related to a submission in a cover letter to the Editor. You should also notify the Editor of any related submissions to other publishers, whether for book or periodical publication, that occur while a manuscript is under review by the *APSR* and which would fall within the scope of this policy. The Editor may request copies of related publications.

If your manuscript contains quantitative evidence and analysis, you should describe your procedures in sufficient detail to permit reviewers to understand and evaluate what has been done and, in the event that the article is accepted for publication, to permit other scholars to carry out similar analyses on other data sets. For example, for surveys, at the least, sampling procedures, response rates, and question wordings should be given; you should calculate response rates according to one of the standard formulas given by the American Association for Public Opinion Research, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (Ann Arbor, MI: AAPOR, 2000). This document is available on the Internet at <[http://www.aapor.org/default.asp?page=survey\\_methods/standards\\_and\\_best\\_practices/standard\\_definitions](http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/standard_definitions)>. For experiments, provide full descriptions of experimental protocols, methods of subject recruitment and selection, subject payments and debriefing procedures, and so on. Articles should be self-contained, so you should not simply refer read-

ers to other publications for descriptions of these basic research procedures.

Please indicate variables included in statistical analyses by capitalizing the first letter in the variable name and italicizing the entire variable name the first time each is mentioned in the text. You should also use the same names for variables in text and tables and, wherever possible, should avoid the use of acronyms and computer abbreviations when discussing variables in the text. All variables appearing in tables should have been mentioned in the text and the reason for their inclusion discussed.

As part of the review process, you may be asked to submit additional documentation if procedures are not sufficiently clear; the review process works most efficiently if such information is given in the initial submission. If you advise readers that additional information is available, you should submit printed copies of that information with the manuscript. If the amount of this supplementary information is extensive, please inquire about alternate procedures.

The *APSR* uses a double-blind review process. You should follow the guidelines for preparing anonymous copies in the Specific Procedures section below.

Manuscripts that are largely or entirely critiques or commentaries on previously published *APSR* articles will be reviewed using the same general procedures as for other manuscripts, with one exception. In addition to the usual number of reviewers, such manuscripts will also be sent to the scholar(s) whose work is being criticized, in the same anonymous form that they are sent to reviewers. Comments from the original author(s) to the Editor will be invited as a supplement to the advice of reviewers. This notice to the original author(s) is intended (1) to encourage review of the details of analyses or research procedures that might escape the notice of disinterested reviewers; (2) to enable prompt publication of critiques by supplying criticized authors with early notice of their existence and, therefore, more adequate time to reply; and (3) as a courtesy to criticized authors. If you submit such a manuscript, you should therefore send as many additional copies of their manuscripts as will be required for this purpose.

Manuscripts being submitted for publication should be sent to Lee Sigelman, Editor, *American Political Science Review*, Department of Political Science, The George Washington University, Washington, DC 20052. Correspondence concerning manuscripts under review may be sent to the same address or e-mailed to [apsr@gwu.edu](mailto:apsr@gwu.edu).

### Manuscript Formatting

Manuscripts should not be longer than 45 pages including text, all tables and figures, notes, references, and appendices. This page size guideline is based on the U.S. standard 8.5 × 11-inch paper; if you are submitting a manuscript printed on longer paper, you must adjust accordingly. The font size must be at least 11 points for all parts of the paper, including notes and references. The entire paper, including notes and references, must be double-spaced, with the sole exception of tables

for which double-spacing would require a second page otherwise not needed. All pages should be numbered in one sequence, and text should be formatted using a normal single column no wider than 6.5 inches, as is typical for manuscripts (rather than the double-column format of the published version of the *APSR*), and printed on one side of the page only. Include an abstract of no more than 150 words. The *APSR* style of embedded citations should be used, and there must be a separate list of references at the end of the manuscript. Do not use notes for simple citations. These specifications are designed to make it easier for reviewers to read and evaluate papers. Papers not adhering to these guidelines are subject to being rejected without review.

For submission and review purposes, you may place footnotes at the bottom of the pages instead of using endnotes, and you may locate tables and figures (on separate pages and only one to a page) approximately where they fall in the text. However, manuscripts accepted for publication must be submitted with endnotes, and with tables and figures on separate pages at the back of the manuscript with standard indications of text placement, e.g., [Table 3 about here]. In deciding how to format your initial submission, please consider the necessity of making these changes if your paper is accepted. If your paper is accepted for publication, you will also be required to submit camera-ready copy of graphs or other types of figures. Instructions will be provided.

For specific formatting style of citations and references, please refer to articles in the most recent issue of the *APSR*. For unusual style or formatting issues, you should consult the latest edition of *The Chicago Manual of Style*. For review purposes, citations and references need not be in specific *APSR* format, although some generally accepted format should be used, and all citation and reference information should be provided.

## Specific Procedures

Please follow these specific procedures for submission:

1. You are invited to submit a list of scholars who would be appropriate reviewers of your manuscript. The Editor will refer to this list in selecting reviewers, though there obviously can be no guarantee that those you suggest will actually be chosen. Do not list anyone who has already commented on your paper or an earlier version of it, or any of your current or recent collaborators, institutional colleagues, mentors, students, or close friends.
2. Submit five copies of manuscripts *and* a diskette or CD containing a pdf file of the anonymous version of the manuscript. If you cannot save the manuscript as a pdf, just send in the diskette or CD with the word-processed version. Please ensure that the paper and diskette or CD versions you submit are identical; the diskette or CD version should be of the **anonymous**

copy (see below). Please review all pages of all copies to make sure that all copies contain all tables, figures, appendices, and bibliography mentioned in the manuscript and that all pages are legible. Label the diskette or CD clearly with the (first) author's name and the title of the manuscript (in abridged form if need be), and identify the word processing program and operating system. If you are unable to create a diskette or CD, please note this in your submission, and you will be asked to e-mail the appropriate file.

3. To comply with the *APSR*'s procedure of double-blind peer reviews, only one of the five copies submitted should be fully identified as to authorship and four should be in anonymous format.
4. For anonymous copies, if it is important to the development of the paper that your previous publications be cited, please do this in a way that does not make the authorship of the submitted paper obvious. This is usually most easily accomplished by referring to yourself in the third person and including normal references to the work cited in the list of references. In no circumstances should your prior publications be included in the bibliography in their normal alphabetical location but with your name deleted. Assuming that text references to your previous work are in the third person, you should include full citations as usual in the bibliography. Please discuss the use of other procedures to render manuscripts anonymous with the Editor prior to submission. You should not thank colleagues in notes or elsewhere in the body of the paper or mention institution names, web page addresses, or other potentially identifying information. *All acknowledgments must appear on the title page of the identified copy only.* Manuscripts that are judged not anonymous will not be reviewed.
5. The first page of the four anonymous copies should contain only the title and an abstract of no more than 150 words. The first page of the identified copy should contain (a) the name, academic rank, institutional affiliation, and contact information (mailing address, telephone, fax, e-mail address) for all authors; (b) in the case of multiple authors, an indication of the author who will receive correspondence; (c) any relevant citations to your previous work that have been omitted from the anonymous copies; and (d) acknowledgments, including the names of anyone who has provided comments on the manuscript. If the identified copy contains any unique references or is worded differently in any way, please mark this copy with "Contains author citations" at the top of the first page.

No copies of submitted manuscripts can be returned.

## ELECTRONIC ACCESS TO THE APSR

Back issues of the *APSR* are available in several electronic formats and through several vendors. Except for the last three years (as an annually "moving wall"), back issues of the *APSR* beginning with Volume 1, Number 1 (November 1906), are available on-line through JSTOR (<http://www.jstor.org/>). At present, JSTOR's complete journal collection is available only via institutional subscription, e.g., through many college and university libraries. For APSA members who do not have access to an institutional subscription to JSTOR, individual subscriptions to its *APSR* content are available. Please contact Member Services at APSA for further information, including annual subscription fees.

Individual members of the American Political Science Association can access recent issues of the *APSR* and *PS* through the APSA website ([www.apsanet.org](http://www.apsanet.org)) with their username and password. Individual non-member access to the online edition will also be available, but only through institutions that hold either a print-plus-electronic subscription or an electronic-only subscription, provided the institution has registered and activated its online subscription.

Full text access to current issues of both the *APSR* and *PS* is also available on-line by library subscription from a number of database vendors. Currently, these include University Microfilms Inc. (UMI) (via its CD-ROMs General Periodicals Online and Social Science Index and the on-line database ProQuest Direct), On-line Computer Library Center (OCLC) (through its on-line database First Search as well as on CD-ROMs and magnetic tape), and the Information Access Company (IAC) (through its products Expanded Academic Index, InfoTrac, and several on-line services [see below]). Others may be added from time to time.

The *APSR* is also available on databases through six online services: Datastar (Datastar), Business Library (Dow Jones), Cognito (IAC), Encarta Online Library (IAC), IAC Business (Dialog), and Newsearch (Dialog).

The editorial office of the *APSR* is not involved in the subscription process to either JSTOR for back issues or the other vendors for current issues. Please contact APSA, your reference librarian, or the database vendor for further information about availability.

## BOOK REVIEWS

The *APSR* no longer contains book reviews. As of 2003, book reviews have moved to *Perspectives on Politics*. All books for review should be sent to the *Perspectives on Politics* Book Review Editor, Jeffrey C. Isaac. The address is Professor Jeffrey C. Isaac, Review Editor, *Perspectives on Politics*, Department of Political Science, Woodburn Hall, 1100 E. 7th St., Indiana University, Bloomington, IN 47405-7110. E-mail: [isaac@indiana.edu](mailto:isaac@indiana.edu).

If you are the author of a book you wish to be considered for review, please ask your publisher to

send a copy to the *Perspectives on Politics* Book Review Editor per the mailing instructions above. If you are interested in reviewing books for *Perspectives on Politics*, please send your vita to the Book Review Editor; you should not ask to review a specific book.

## OTHER CORRESPONDENCE

The American Political Science Association's address, telephone, and fax are 1527 New Hampshire Avenue, N.W., Washington, D.C. 20036, (202) 483-2512 (voice), and (202) 483-2657 (fax). E-mail: [apsa@apsanet.org](mailto:apsa@apsanet.org). Please direct correspondence as follows.

Information, including news and notes, for *PS*:

Dr. Robert J-P. Hauck, Editor, *PS*  
E-mail: [rhauck@apsanet.org](mailto:rhauck@apsanet.org)

Circulation and subscription correspondence (domestic claims for nonreceipt of issues must be made within four months of the month of publication; overseas claims, within eight months):

Sean Twombly,  
Director of Member Services  
E-mail: [membership@apsanet.org](mailto:membership@apsanet.org)

Reprint permissions:  
E-mail: [Rights@cambridge.org](mailto:Rights@cambridge.org)

Advertising information and rates:

Advertising Coordinator,  
Cambridge University Press  
E-mail: [advertising@apsanet.org](mailto:advertising@apsanet.org)

## EXPEDITING REQUESTS FOR COPYING APSR AND PS ARTICLES FOR CLASS USE AND OTHER PURPOSES

### Class Use

The Comprehensive Publisher Photocopy Agreement between APSA and the Copyright Clearance Center (CCC) permits bookstores and copy centers to receive expedited clearance to copy articles from the *APSR* and *PS* in compliance with the Association's policies and applicable fees. The general fee for articles is 75 cents per copy. However, current Association policy levies no fee for the first 10 copies of a printed article, whether in course packs or on reserve. Smaller classes that rely heavily on articles (i.e., upper-level undergraduate and graduate classes) can take advantage of this provision, and faculty ordering 10 or fewer course packs should bring it to the attention of course pack providers. APSA policy also permits free use of the electronic library reserve, with no limit on the number

of students who can access the electronic reserve. Both large and small classes that rely on these articles can take advantage of this provision. The CCC's address, telephone, and fax are 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400 (voice), and (978) 750-4474 (fax). This agreement pertains only to the reproduction and distribution of APSA materials as hard copies (e.g., photocopies, microfilm, and microfiche).

The Association of American Publishers (AAP) has created a standardized form for college faculty to submit to a copy center or bookstore to request copyrighted material for course packs. The form is available through the CCC, which will handle copyright permissions.

APSA also has a separate agreement pertaining to CCC's Academic E-Reserve Service. This agreement allows electronic access for students and instructors of a designated class at a designated institution for a specified article or set of articles in electronic format. Access is by password for the duration of a class.

Please contact your librarian, the CCC, or the APSA Reprints Department for further information.

### **APSR Authors**

If you are the author of an *APSR* article, you may use your article in course packs or other printed materials without payment of royalty fees and you may post it at personal or institutional web sites as long as the APSA copyright notice is included.

### **Other Uses of APSA-Copyrighted Materials**

For any further copyright issues, please contact the APSA Reprints Department.

### **INDEXING**

Articles appearing in the *APSR* before June 1953 were indexed in *The Reader's Guide to Periodical Literature*. Current issues are indexed in *ABC Pol Sci; America, History and Life 1954-*; Book Review Index; Current Contents: Social and Behavioral Sciences; Econ-Lit; *Energy Information Abstracts*; Environmental Abstracts; Historical Abstracts; Index of Economic Articles; Information Service Bulletin; International Index; International Political Science Abstracts; the *Journal of Economic Literature*; Periodical Abstracts; Public Affairs; Public Affairs Information Service International Recently Published Articles; Reference Sources; Social Sciences and Humanities Index; Social Sciences Index; Social Work Research and Abstracts; and Writings on American History. Some of these sources may be available in electronic form through local public or educational libraries. Microfilm of the *APSR*, beginning with Volume 1, and the index of the *APSR* through 1969 are available through University Microfilms Inc., 300 North Zeeb Road, Ann Arbor, MI 48106 ([www.umi.com](http://www.umi.com)). The Cumulative Index to the *American Political Science Review*, Volumes 63 to 89: 1969-95, is available through the APSA.

# Majority Norms, Multiculturalism, and Gender Equality

SARAH SONG *Massachusetts Institute of Technology*

**A**lthough many scholars have discussed the conflict that can arise between multiculturalism and gender equality, both critics and defenders of multiculturalism have largely overlooked a variety of interactive dynamics between majority and minority cultures that have important implications for the theory and practice of multiculturalism. Examining cases in the U.S. context, this essay argues for an interactive view of the dilemmas of gender and culture that is attentive to interconnections between majority and minority cultures. What is of particular concern for debates on multiculturalism is that the mainstream legal and normative frameworks within which minority claims for accommodation are evaluated have themselves been informed by patriarchal norms, which in turn have offered support for gender hierarchies within minority cultures. The interactive view defended here suggests the need to scrutinize both minority and majority norms and practices in evaluating the claims of minority cultures.

**I**n 1975, Julia and Audrey Martinez filed suit against the Santa Clara Pueblo. Julia Martinez was a Pueblo woman who had married a man outside the tribe, and although her daughter Audrey had been raised within the Pueblo community, she was denied membership in the tribe. According to tribal membership rules, instituted in 1939, women who married out of the tribe could not transmit their membership to their children, whereas men who married out could pass their membership to their children. At stake for Audrey Martinez was not only recognition as a tribal member but also the political rights and material benefits of tribal membership, including health care, education, and housing assistance from the federal government. On appeal, the U.S. Supreme Court ruled that it could not hear the equal protection claim on the grounds that it did not have jurisdiction over matters of tribal membership. If the federal courts were to intervene in tribal decisions, the Court added, they would interfere with the "tribe's ability to maintain itself as a culturally and politically distinct entity" (*Santa Clara Pueblo v. Martinez*, 436 U.S. 49 [1978], 123–24).

In 1988, a Chinese immigrant, Dong Lu Chen, who had lived in New York City for 1 year, discovered that his wife was having an affair. A few weeks later he beat and killed her. Chen confessed that he killed his wife because she had committed adultery. An anthropologist testified that violence against unfaithful spouses

was commonplace in Chinese culture. Chen was found guilty of second-degree manslaughter. Relying heavily on cultural evidence, the judge sentenced him to 5 years' probation with no jail time, a much lighter punishment than that usually associated with a second-degree manslaughter conviction (*People v. Chen*, No. 87-7774, Supreme Court, NY County [Dec. 2, 1988]).

These two cases illustrate the conflict that can arise between multiculturalism and gender equality—that is, between granting special accommodations to cultural minorities and pursuing equality between the sexes. Multiculturalism has been defended on a number of grounds. According to the prominent "liberal culturalist" position developed by Will Kymlicka (1989, 166; 1995, 83), cultures are an important good; they serve as "contexts of choice." Individual freedom requires having options from which to choose, and it is cultures that provide and give meaning to these options. But members of minority cultural communities face special disadvantages with regard to cultural membership because states tend to establish one language and culture, usually the majority's, as the public norm. Kymlicka (1989, 1995, 2001) argues that treating members of minority cultural groups as equals requires special accommodations to protect their contexts of choice. Although group-differentiated treatment has been defended for women and racial minority groups, *cultural* groups—including immigrants, religious minorities, and indigenous peoples—have been the primary focus of recent debates about group-differentiated citizenship. Different types of cultural accommodations include consideration of one's group membership in the application of the law, exemptions to generally applicable laws, group-specific family law, and self-government rights (Levy 1997).<sup>1</sup>

Sarah Song is Assistant Professor, Department of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E53-435, Cambridge, MA 02139 (ssong@mit.edu).

Earlier versions of this paper were presented at the 2004 meeting of the Democracy Collaborative at the University of Maryland, the 2004 meeting of the American Political Science Association in Chicago, the Brandeis Colloquium on Democracy and Cultural Pluralism, and the Harvard Political Theory Colloquium. I am grateful to all participants at those meetings, especially Stephen Elkin, Mark Graber, Jane Mansbridge, Nancy Rosenblum, Ayelet Shachar, Marion Smiley, Robin West, and Bernard Yack. I am also grateful for comments from Seyla Benhabib, Suzanne Berger, Kanchan Chandra, Joshua Cohen, Sally Haslanger, Vicki Hsueh, Theresa Lee, Tamara Metz, Melissa Nobles, Susan Okin, Jennifer Pitts, Ian Shapiro, Lee Sigelman, Rogers Smith, Verity Smith, and three anonymous reviewers.

<sup>1</sup> I use the terms *multiculturalism* and *cultural accommodation* over *cultural rights* because the former terms encompass both rights-based and non-rights-based accounts of special protections for cultural minorities. Non-rights-based accounts aim to protect what are taken to be important interests and values of cultural minorities without expressing those interests and values in terms of rights. The question of whether cultural protections should be understood as rights is an important question, but one I do not seek to resolve here.



Yet, as many political theorists starting with Susan Okin have emphasized, multiculturalism can have the effect of reinforcing gender inequality within the minority groups being accommodated (e.g., Deveau 2000, 2003; Okin 1998, 1999; Shachar 2001). The idea here is that by granting accommodations to a minority cultural group, states permit some members—usually the group's more powerful members—to oppress more vulnerable members within the group. The problem of vulnerable members within minority groups, or “internal minorities,” may apply not only to women but also to religious, linguistic, and sexual minorities (Green 1995). Feminists have been most concerned with the effects of cultural accommodation on women. For instance, Okin focuses on “cultural defense” cases involving immigrants and religious communities in Western societies, and Ayelet Shachar examines legal pluralist institutional arrangements that grant religious groups autonomy over family law, as in Israel and India, to illustrate how cultural accommodation can work to the detriment of women.

But there is another dimension to the practice of cultural accommodation that may be, in Okin's (1999) words, “bad for women” that both defenders and critics of multiculturalism have largely overlooked. Debates about multiculturalism seem largely to have assumed that the obstacles to improving the status of minority women have to do with the gender norms and practices of minority cultures. Call this the *internal view*. To be sure, the problem of gender inequality within minority communities stems in part from struggles internal to the culture, but this internal view overlooks the ways in which gender statuses are shaped by intercultural interactions.

This essay examines a range of cases in the U.S. context with two aims in mind: first, to propose an alternative *interactive view* of the dilemmas of gender and culture that is attentive to how majority and minority cultures interact in hierarchy-reinforcing ways, sometimes through the practice of multiculturalism, and second, to discuss some of the implications of adopting this interactive view for the theory and practice of multiculturalism. Minority norms and practices can threaten the pursuit of gender equality within the majority culture, but influence also runs in the other direction. Majority norms and practices also pose obstacles to the pursuit of gender equality within minority cultures. What is of particular concern for the debate on multiculturalism is that the mainstream legal and normative frameworks within which minority claims for accommodation are evaluated have themselves been informed by patriarchal norms, which in turn have offered support for gender hierarchies within minority cultures. This suggests the need to be attentive to both minority and majority norms and practices in evaluating the claims of minority cultures.

## CONCEPTUALIZING CULTURE

Before turning to the cases, I want to consider the way culture has been conceptualized in the debate on

multiculturalism, which I believe prevents us from adequately capturing the complexity of cultural conflicts. Much of the debate on multiculturalism has focused on normative claims about culture, including whether cultural minorities are entitled to special treatment and to what types of treatment. Much less explicit attention has been given to the way culture gets conceptualized. Take Kymlicka's influential defense of multiculturalism. It rests on a view of cultures as well-integrated, well-bounded, and largely self-generated entities, defined by a set of key attributes, including a shared language, history, and values. These key attributes are institutionally embodied in schools, the media, the economy, and government. The sort of culture Kymlicka seeks to protect is what he calls “societal culture,” which “provides its members with meaningful ways of life across the full range of human activities, including social, educational, religious, recreational, and economic life, encompassing both public and private spheres. These cultures tend to be territorially concentrated, and based on a shared language” (Kymlicka 1995, 76). In Kymlicka's (1995, 80) view, societal cultures tend to be national cultures, which are pervasive or encompassing.

But cultures may not be as unified or well-bounded as Kymlicka's view suggests. Cultures are more realistically conceived as frameworks of meaning shaped and reshaped by the words and actions of their members (Benhabib 2002; Ortner 1984, 1996; Wedeen 2002). On this social constructivist view, cultural identity is understood as one of many socially constructed categories in which membership is based on certain publicly identifiable attributes. The criteria of ascription associated with cultural identity groups include shared ancestry (actual or mythical), a common language, and shared customs. But there is also a subjective dimension: to what extent individuals embrace the categories and what meanings individuals attach to them (Appiah 2004, 21–23). One might possess attributes associated with particular categories but not experience a strong sense of belonging or attachment to them; indeed, one may resist and seek to transform the categories themselves. Cultures and cultural identities emerge, change, and are maintained through social interactions and political struggle. To say that they are constructed in and through social interactions is not to say that they are false or weak, nor that they are always radically in flux. But neither are cultures natural facts about the world. To get at cultural differences, we need to examine the processes by which cultural identities get constructed. This requires attention to historical context and to collective discussions and struggles within a community, as well as to interactions between members and outsiders—in particular, to the role of states in shaping cultural identities.

Kymlicka's (1995) view of culture pays insufficient attention to the *politics* of cultural construction and conflict and thereby downplays the extent to which cultures are internally varied and contested (Benhabib 2002, 25, 65; Parekh 2000, 157, 175; Shachar 2001, 3). He brackets the politics of cultural construction and conflict by distinguishing between a culture's

"structure" and its "character"; the former consists of what is essential for the culture's survival, whereas the latter includes particular norms and practices. A culture's structure is "more fundamental" than a culture's character because it is the former that provides the "context of choice" for its members, and so it is the culture's structure, not its particular character or norms, that is defended in his account (1989, 167–68, 172; 1995, 104–105). But this distinction between structure and content is itself the product of political struggle. What is seen as essential for the survival of a culture such that it is deemed a part of the "cultural structure" is itself the result of a community's struggles over meaning and power. For instance, the leaders of a minority cultural community might argue, against Kymlicka, that their cultural community continues to exist precisely because its members continue particular patriarchal practices, and that if such practices are not protected, then the culture will become extinct. In other words, they might argue that their right to cultural context is a right to particular cultural content (Margalit and Halbertal 1994, 504–505). Kymlicka himself suggests a definition of culture that is more than the mere existence of a cultural community and includes some particular content. In his early work, he defines culture in terms of history, language, and traditions, and in his later work, he defines "societal culture" as "a meaningful way of life" based on a common language and common values (1989, 135, 165; 1995, 76). This move of defining culture in terms of particular content opens the door to its own set of challenges, not least determining which practices are indeed central for the maintenance of a culture, and how the centrality of a particular practice should be weighed against other concerns. But these are not matters that can be resolved once and for all by definitional fiat; it is a matter to be decided through the process of collective discussion and meaning-making.

In addition, Kymlicka's view of culture understates the extent to which cultures are interactive and thus overlooks the ways in which the content of a culture and its change are shaped by other cultures and not only by internal conflict. Cultures have long interacted and mutually influenced one another through relations of trade, warfare, and conquest, and cultures do not correspond in any neat way to national or societal boundaries. When it comes to minority cultures within one state, the influence of the dominant culture is undeniable. Jeremy Waldron has emphasized the interactive nature of cultures in criticizing Kymlicka's strategy of cultural preservation, which he views to be based on a flawed understanding of the nature of cultures: "To *preserve* a culture is often to take a favored 'snapshot' of it, and insist that this version must persist at all costs, in its defined purity, irrespective of the surrounding social, economic, and political circumstances." But such a strategy of preservation would "cripple the mechanisms of adaptation and compromise," which is an "inherent feature" of every culture (Waldron 1995, 109–110). In his reply to Waldron, Kymlicka does recognize the fact of cultural interchange; indeed, one might argue that Kymlicka thinks

there is too much interchange from majority to minority cultures such that the latter are threatened with extinction. But he has focused primarily on one type of interaction—Western states' strategy of "benign neglect" toward minority cultures—in order to expose what he deems to be the incoherence and injustice of this strategy. In practice, however, indifference is not the only or even primary mode of interaction between Western states and minority groups. The dominant culture's own unjust norms have shaped the practice of cultural accommodation, leading in some cases to the accommodation of unjust practices within minority cultures. In addition, although we might agree with Kymlicka's (1995, 105) claim against Waldron that acknowledging cultural interchange does not mean accepting that there are no distinct cultures, we need to recognize more than Kymlicka does that the distinctness of a culture depends in part on the nature and extent of cultural interchange. Rather than assuming that cultures are distinct and largely endogenously developed wholes, we should instead be open to questioning to what extent they are.

Several recent contributions to the debate on multiculturalism have emphasized that cultural interactions—through the global economy, transnational communications networks, and migrations of people across borders—are an important source of cultural change. As Bhikhu Parekh (2000, 163) puts it, "[C]ultures are not the achievements of the relevant communities alone but also of others, who provide their context, shape some of their beliefs and practices, and remain their points of reference. In this sense almost all cultures are multiculturally constituted." Both Benhabib (2002, 7) and Deveau (2003, 790) have emphasized the permeability of boundaries between cultures. Shachar (2001, 2, 88–92, 117–45) stresses that groups are always reacting to the effects of state power, and her "joint governance" approach, which calls for ongoing interaction between the state and minority groups in the governance of different spheres of minority group life, clearly recognizes that minority and majority cultures are interconnected.<sup>2</sup>

Although these theorists have stressed that cultural interactions are an important source of cultural construction and change, they stop short of examining how cultural interactions have shaped cultural identities and conflicts. A social constructivist view of

<sup>2</sup> Shachar defines "identity groups" or "nomoi communities" as "religiously defined groups of people" who "share a comprehensive and distinguishable worldview that extends to creating a law for the community." Identity groups are said to share "a unique history and collective memory, a distinct culture, a set of social norms, customs, and traditions" (Shachar 2001, 2, n. 5). But this definition seems to "recapitulate the mistakes of group essentialism" (Benhabib 2002, 123) because members of cultural groups may not share a comprehensive worldview and because cultures are not endogenously developed wholes as Shachar's definition of "identity groups" seems to suggest. Although Shachar develops institutional designs aimed at promoting interaction between states and minority groups in the governance of minority affairs, her analysis does not examine the role that states have played in shaping and reinforcing minority group identities and practices at the center of cultural conflicts.

cultures recognizes that many cultural conflicts may indeed be intracultural; for instance, conflicts over female circumcision or customary marriage may primarily be struggles within particular cultural communities over the meaning and importance of particular practices (Deveaux 2003, 784). But many cultural conflicts arise out of intercultural interactions, and what appear to be intracultural conflicts may have been fueled by intercultural interactions. In some cases, intercultural interactions may provoke hardening of hierarchies within minority groups, as in cases where group leaders shore up traditional decision-making structures within the community in the face of external challenges to those structures. In other cases, majority institutions may directly or indirectly support gender hierarchies within minority communities. Cultures vary in the degree of fluidity, contestation, and permeability, but even in relatively closed groups the content of cultures is not determined purely "from the inside." Rather, cultures have developed through interactions and struggles with other cultures. This suggests the need to examine more closely the role of interactive dynamics in shaping the identities and practices of majority and minority cultural communities.

### INTERCONNECTIONS BETWEEN MAJORITY AND MINORITY CULTURES

There are a great variety of ways that cultures have interacted; my focus here is on a range of interactive dynamics that have had the effect of reinforcing gender hierarchies across cultures. To be sure, cultural interchange may push toward greater gender equality within cultures, but I limit my focus to hierarchy-reinforcing interactions because they have been given less attention in debates on multiculturalism.

Majority cultures have long shaped the gender practices of minority cultures. For instance, majority institutions directly imposed mainstream gender biases onto minority cultural communities, as in the case of the 1887 Dawes Act, which subverted Native American women's roles in agricultural work by making Native American men heads of household, landowners, and farmers (Cott 2000, 123). More common today and of greater concern for contemporary debates on multiculturalism are the indirect ways in which mainstream gender norms have resonated with and offered support for gender hierarchies in minority cultural communities, as in the cases of the Santa Clara Pueblo and the "cultural defense." I call this type of interaction the *congruence effect*. Influence can also run in the other direction with minority norms shaping the gender practices of majority cultural communities: the accommodation of patriarchal practices within minority cultural communities may feed back and reinforce gender inequality within the wider society (*boomerang effect*), and even in cases where minority claims for accommodation are denied, as in the case of Mormon polygamy or female circumcision, the focus on the patriarchal practices of minority cultures can have the effect of diverting attention from gender hierarchies within the majority culture (*diversionary effect*).

### Majority → Minority: Congruence Effect and the Case of the Santa Clara Pueblo

Consider first the majority culture's influence on minority cultures. The struggle toward gender equality in the majority culture is incomplete and ongoing, and in some cases, sexist practices in minority cultures have been congruent with and received support from the majority culture's own gender-biased traditions. The Santa Clara Pueblo case I briefly discussed at the outset is a good example. Julia Martinez and her daughter tried to persuade the tribe to change its gender-biased membership rules, and when their efforts met without success, they filed a lawsuit under the Indian Civil Rights Act of 1968. The core issue in this case was not whether the Pueblo had denied equal protection for out-marrying women (this was recognized to be the case), but rather, what the limits of tribal sovereignty should be—that is, whether a minority group's right to define its own membership rules should be allowed to take priority over ensuring equal protection for women.

The U.S. District Court for the District of New Mexico ruled in favor of the tribal government, arguing that the membership rule reflected deep-seated patriarchal traditions of the tribe and that undermining tribal decisions over membership rules would destroy Pueblo culture. The District Court argued that the male–female distinction was "rooted in certain traditional values," the Pueblo's patrilineal and patrilocal traditions, and concluded, "To abrogate tribal decisions, particularly in the delicate area of membership, for whatever 'good' reasons, is to destroy cultural identity under the guise of saving it" (*Martinez v. Santa Clara Pueblo*, 402 F. Supp. 5 [1975], 16, 19). On appeal, the Court of Appeals for the Tenth Circuit overturned the District Court ruling in part because it rejected the view of Pueblo culture as homogenous and generally patriarchal. The court acknowledged that the tribe had an interest in retaining its culture: "[W]here the tribal tradition is deep-seated and the individual injury is relatively insignificant, courts should be and have been reluctant to order the tribal authority to give way" (*Martinez v. Santa Clara Pueblo*, 540 F.2d 1039 [1976], 1047). But rather than taking the cultural claim at face value, the appellate court scrutinized the cultural traditions of the Pueblo, questioning to what extent the gender-biased membership rule was integral to Pueblo culture. The court ruled that the tribal interest in upholding the membership rule was not substantial enough to justify its discriminatory effect on the grounds that the membership rule was not a part of long-standing Pueblo tradition but motivated by "economics and pragmatics," and that it did not rationally identify those persons who were culturally Santa Clarans—the Martinez children had grown up with the Pueblo, spoke the language of the Pueblo, and practiced the Pueblo's religion and customs (1048). The Santa Clara Pueblo appealed to the Supreme Court.

Writing for the majority, Justice Thurgood Marshall did not address the equal protection issue involving the charge of gender discrimination, limiting the Court's consideration to the question of federal review

of tribal policy (*Santa Clara Pueblo v. Martinez*, 436 U.S. 49 [1978]). Although he acknowledged that Indian nations possess separate sovereignty that pre-existed the U.S. Constitution and thus fall beyond its constraints, Marshall reasoned that the traditional powers of Indian self-government could be modified or eliminated by congressional enactment. Congress then, through its plenary power, could lawfully pass the Indian Civil Rights Act (ICRA) imposing certain restraints on tribal governments. However, the ICRA imposed "certain restrictions upon tribal governments similar, but not identical, to those contained in the Bill of Rights and the Fourteenth Amendment"; the only express appeal remedy that Congress provided in the ICRA was the writ of habeas corpus, which did not help the Martinez children because their case did not involve detention by the tribe (63). Although the Court focused on the procedural question of federal review of tribal policy, it went beyond purely procedural considerations by linking the question of tribal jurisdiction with a substantive concern for the maintenance of tribal identity. It argued that if the federal courts were to intervene in tribal decisions they "may substantially interfere with the tribe's ability to maintain itself as a culturally and politically distinct entity" (72). The Court concluded that no cause of action existed for equal protection claims, such as the one raised by Julia and Audrey Martinez, and therefore, the federal courts could not hear the discrimination charge.

Feminist critics have pointed to this case as an example illustrating the conflict between multiculturalism and gender equality, but what they have overlooked is the role the U.S. government played in the creation of the Santa Clara Pueblo membership rule. Far from being foreign or different, the Pueblo's gender-biased amendment to the membership rule was congruent with the majority culture's own norms and policies on membership. In 1935, the U.S. Secretary of Interior approved the Santa Clara Pueblo's Constitution and Bylaws, which extended membership to four groups of people: those "of Indian blood" whose names appeared on the 1935 census roll; all persons born of parents who are both members of the tribe; all "children of mixed marriages between members of the Santa Clara Pueblo and nonmembers, provided such children have been recognized and adopted by the council"; and all "persons naturalized as members of the Pueblo" (Brief of the Petitioners, *Santa Clara Pueblo v. Martinez*, No. 76-682 Appendix 1). In 1939, the Pueblo amended its membership rules, stating: "[C]hildren born of marriages between female members of the Santa Clara Pueblo and nonmembers shall not be members," and "[p]ersons shall not be naturalized as members of the Santa Clara Pueblo under any circumstances" (18). Only two groups were eligible for membership: children born of marriages between members of the Santa Clara Pueblo, and children born of marriages between male members of the Santa Clara Pueblo and female nonmembers.

It does not appear that the U.S. government directly mandated or suggested membership restrictions along

gender lines to the Santa Clara Pueblo. Although the kinship systems of the Pueblos had traditionally been organized along matrilineal lines, by the turn of the twentieth century, the Eastern Pueblos, including the Santa Clara Pueblo, were no longer organized straightforwardly in matrilineal terms. Partly as a result of the efforts of Spanish colonists and Franciscan friars to break down the Pueblos' matrilineal kinship patterns, by the late nineteenth century, the Eastern Pueblos no longer organized matrilineally (Jacobs 1999, 7). At trial, one anthropologist testified that the Pueblo had a patrilineal kinship system, suggesting that the 1939 membership restriction grew out of Pueblo traditions (Brief of the Respondents, *Santa Clara Pueblo v. Martinez*, 540 F.2d 1039 [1976], 36-37). Anthropologist and Santa Clara Pueblo member, Edward Dozier (1970, 133-34, 145-48, 163-66), also maintained that the Santa Clara Pueblo no longer organized matrilineally at the time of the 1939 membership rule change, though he argued that the Eastern Pueblos had bilateral kinship systems in which lineage was determined by both parents.<sup>3</sup>

What the U.S. government did do was to pressure the Pueblo and other tribes to adopt more restrictive membership rules, and it reviewed and approved the Pueblo membership restriction along gender lines. As Audra Simpson has observed in the context of studying narratives of citizenship among Kahnawake Mohawks, Native peoples "witness the forced cultural transformation of native culture through the bounding of people and bounding space" (Simpson 2000, 118; see also Resnik 1989, 719). State efforts at boundedness are represented in the creation of reservations and compulsion to restrict membership. The idea of "membership" was itself imposed by the U.S. government in order to count Native peoples and regulate the resources it distributed to tribes (Resnik 1989, 719-22). The Pueblo tribal authority moved to restrict its membership in direct response to a federal government circular. On November 18, 1935, in a circular titled "Membership in Indian Tribes," the U.S. Department of Interior made

<sup>3</sup> It is important to note that there is no straightforward relationship between mother/father-based lineage systems and male/female power in a society. Anthropological accounts of this relationship among the Pueblo may say more about the ideology and concepts of anthropologists than they do about Pueblo norms and practices (Green 1980). One anthropologist finds that the pattern of house inheritance among the Santa Clara Pueblo was "prevalingly patrilineal," and argues women were considered "second-class citizens at Santa Clara Pueblo," suggesting a strong connection between patrilineality and patriarchy (Hill 1982, 20, 169). In contrast, another anthropologist, Elsie Clew Parsons, detailed a less rigid sexual division of labor among the Pueblo in which men and women reversed roles on some occasions, leading her to conclude that sex-based role differentiation "appear[ed] to count very little if at all in their personal relations, but in their occupations it is all controlling" (quoted in Jacobs 1999, 6). Parson's account of the Pueblo must be read in light of her attempts to use her study of the Pueblo as a way to articulate an alternative ideal of gender relations for the dominant culture—one in which sex differentiation counted for very little and in which there was a healthy attitude toward sexuality (Jacobs, 74, 78). Although the revised Pueblo membership rule may not have reflected deeply rooted patriarchal traditions among the Pueblo, it did reflect a gender bias, which one Pueblo woman and her daughter sought to contest.

a “declaration” to all “engaged in Indian Reorganization Act” that “Congress [has] a definite policy to limit the application of Indian benefits” (Circular No. 3123 (Office of Indian Affairs, Nov. 18, 1935), in *Opinions of the Solicitor General* [April 12, 1938]). The Department planned “to urge and insist that any constitutional provision conferring automatic tribal membership upon children hereafter born, should limit such membership to persons who reasonably can be expected to participate in tribal relations and affairs.” The government suggested ways to restrict membership, which reflected its own views about the proper bases of political membership—in particular, that both parents be recognized as tribal members or that an individual possess a “certain degree of Indian blood,” as opposed to a shared language or history.

Although the U.S. government did not impose gender-biased membership restrictions onto the Pueblo, America’s own tradition of gendered citizenship laws helped legitimate gendered membership rules among the Pueblo. Into the 1930s, American women endangered their citizenship status by marrying foreign men, whereas American men who married foreign women automatically made their wives into U.S. citizens. In 1855, Congress passed a statute that made married women’s citizenship dependent on their husbands’ citizenship. The 1855 Naturalization Act declared, “Any woman who is now or may hereafter be married to a citizen of the United States and who might herself be lawfully naturalized shall be deemed a citizen” (10 Stat. 604, as reenacted in *Revised Statutes of the U.S.* [1878], sect. 1994). Politicians and judges tended to interpret this act as a mandate to assign married women, whether foreign or American, the citizenship of their husbands. Foreign women who married American men automatically became citizens, making such women the first group of adults to receive U.S. citizenship derivatively. The 1855 law also granted American citizenship to children born abroad to American fathers and foreign mothers, but not to children born abroad to American mothers and foreign fathers. By making wives’ and children’s nationality dependent on the male citizen’s, this law affirmed male headship of the family as a political norm and enhanced the citizenship privileges of American men (Bredbenner 1998; Cott 1998).

In contrast, under the 1855 law, American women who married foreign men were largely seen as forfeiting their U.S. citizenship; such outmarrying women were expected to take up the nationality of their husbands. Some federal judges, as well as the State Department, had generally agreed that a female citizen who married an alien resident did not endanger her American citizenship unless she moved permanently to her husband’s country. But other federal judges maintained that American women lost their citizenship simply by marrying foreign men (*Leonard v. Grant*, 5 F. 11 [1880]; *Pequignot v. Detroit*, 16 F. 211 [1883]; see Bredbenner 1998, 57–60; Smith 1997, 389–90). This ambiguity over whether an American woman forfeited her citizenship by marrying a foreign man was clarified by the Expatriation Act of 1907, which made this gender-biased

policy official: “Any American woman who marries a foreigner shall take the nationality of her husband” (Expatriation Act of 1907, 34 Stat. 1228). American women who married foreign men were expatriated, regardless of residency. This law discouraged American women from marrying immigrants and prevented the wives of immigrant couples from being naturalized on their own.

These gender asymmetries in naturalization and citizenship policy were only partially overturned by the Cable Act, or Married Women’s Independent Nationality Act of 1922. An American woman who married a foreigner and remained in the United States would now remain a U.S. citizen, but she would lose her citizenship if she lived in her husband’s country for 2 years or if she married a man “ineligible for citizenship”—an Asian, a polygamist, or an anarchist. In contrast, an American man did not suffer such consequences for similar actions. There was a similar asymmetry for married immigrant couples seeking naturalization in the United States: an immigrant woman’s ability to pursue naturalization or maintain U.S. citizenship continued to depend on her spouse’s eligibility for naturalization. As Nancy Cott puts it, the Cable Act reflected “the reluctance of Congress to give up its long-term priority for the male citizen as family head” (Cott 2000, 165; see also Bredbenner 1998, 97–100). Women’s citizenship continued to depend on their husbands’ citizenship status until legislative reforms in the 1930s. By 1934, American women no longer were seen to forfeit their citizenship by marrying foreigners, both sexes gained the same naturalization benefits for their foreign-born spouses, and mothers gained the same rights as fathers to pass down citizenship to their children born abroad (Cott 1998, 1469). But these acts did not amount to the attainment of full citizenship for married women. By 1934, women had won suffrage and access to political parties and officeholding, but they had not attained full access to the rights of citizenship.<sup>4</sup>

Seen in this larger context—that is, in light of the majority culture’s own traditions in which married women’s political membership was seen to depend on their husbands’—the Santa Clara Pueblo’s gender-biased membership rule appears not foreign but very familiar to the majority culture’s own gender norms at the time. In reviewing and approving the Pueblo membership rule in 1939, the U.S. Secretary of Interior endorsed a gender-biased rule that was congruent with the majority culture’s own gendered traditions of

<sup>4</sup> For example, from the 1920s to 1975, many states resisted equal admission of women to juries. Citizenship law continues to remain gendered in certain respects. See, for example, *Tuan Anh Nguyen v. I.N.S.*, 533 U.S. 53 (2001), which held that the federal law providing different citizenship rules for children born abroad and out of wedlock depending on whether the citizen parent is the mother or father is consistent with the equal protection guarantee in the Fifth Amendment’s Due Process Clause. The ruling seems to reflect the notion that mothers must care for “illegitimate” children, whereas fathers may ignore them. In addition, as Okin has argued (1989), longstanding gendered division of labor within families continues to hinder women’s full inclusion in economic and political life.

membership. This congruence suggests that accommodation was based not so much on respect for cultural differences and concerns about tribal sovereignty but on the congruence of gender-biased traditions across cultures. Tribal sovereignty is not absolute; the federal government has intervened in tribal affairs when it perceives that tribal decisions may undermine concerns that the government deems to be urgent. For example, in the early 1980s, the Secretary of Interior rescinded a tribal ordinance of the Moapa Band of Paiute Indians that would have permitted houses of prostitution on the Moapa Reservation in Nevada. Although the Department of Interior had approved the Moapa Constitution that included provisions for prostitution, the Secretary of Interior retracted the approval on the grounds that such practices were "frowned upon by federal policy" (*Moapa Band of Paiute Indians v. U.S. Department of Interior*, 747 F.2d 563 [1984]; see MacKinnon 1987 and Resnik 1989). In contrast, the gender-biased membership rule of the Santa Clara Pueblo was not "frowned upon by federal policy" in part because the Pueblo rule was congruent with the dominant culture's own gendered understandings of membership. Adequately capturing and addressing the problems raised by the Pueblo case requires recognizing this intercultural congruence and the state's role in underwriting the Pueblo's gender-biased membership practices

### Majority → Minority: Congruence Effect and the Case of the "Cultural Defense"

The "cultural defense" has been raised in many areas of American law in both civil and criminal cases. In civil cases, individuals ask judges to consider cultural traditions in custody battles, in decisions over medical treatment for children, and in employment discrimination cases. In criminal cases, cultural evidence is presented in order to provide insight into the defendant's state of mind. Cultural evidence in criminal cases has been introduced at various stages: before trial to determine whether to arrest and prosecute; during trial to negate an element of a crime or support an established defense, such as consent or provocation; during sentencing to mitigate punishment; or on appeal to overturn convictions on the grounds that the judge improperly excluded cultural evidence or failed to instruct juries properly on the consideration of cultural evidence (see Renteln 2004). If successful, "cultural defenses" may reduce or eliminate criminal charges, as well as mitigate punishment. In some cases, they have been used by immigrant men in defense against charges of violent crimes against women.

In 1984, a 23-year-old Hmong refugee, Kong Pheng Moua, who had lived in the United States for 6 years, abducted a 19-year-old Hmong woman and forced her to have sex with him. The woman, Xeng Xiong, later called the police and accused the defendant of kidnapping and rape (*People v. Moua*, No. 315972-0, Fresno County Superior Court [Feb. 7, 1985]). In his defense, Moua claimed that he was performing a traditional

Hmong practice of matrimony called "marriage by capture" in which even a woman who is willing to get married should resist in order to establish her virtue. Moua did not present cultural evidence to claim that he did not know that rape was illegal in the United States, nor did he argue that rape was not a category of offense in Hmong culture. Instead, he claimed that he did not understand Xiong's resistance as expressing nonconsent. Moua's lawyer stressed that a Hmong woman's resistance is an important part of the Hmong courtship custom: "At the last minute the girl must say, 'No, no, I'm not ready,' and the boy must say, 'Baloney, you'll be mine tonight.' If those attitudes were not expressed, the girl would not appear virtuous enough to the man, and he would not appear strong enough to her" (Sherman 1986, 36). The court dismissed the rape and kidnapping charges, and Moua was charged with false imprisonment and sentenced to 120 days in jail and a \$1,000 fine, \$900 of which was paid to the victim as a form of restitution.

In another case, a Chinese immigrant, Dong Lu Chen, killed his wife after discovering she was having an affair (*People v. Chen*, No. 87-7774, Supreme Court, NY County [Dec. 2, 1988]). In his defense, an anthropologist testified that "in traditional Chinese culture, a wife's adultery is considered proof that a husband has a weak character, making him undesirable even after a divorce," and because of this stigma, a cuckolded man who reacts violently is behaving reasonably. Asked to compare the defendant's reaction to that of a "reasonable" American husband, the anthropologist stated, "In general terms, I think that one could expect a Chinese to react in a much more volatile, violent way to those circumstances than someone from our own society." The anthropologist did not cite any cases where Chinese men had killed adulterous wives, nor did he present any evidence showing that such killings would go unpunished under Chinese law. The prosecution thought the court would deny the use of cultural evidence, and thus neither challenged the expert about the cultural evidence nor raised competing evidence (Sherman 1989, 28). The judge noted that Chen's cultural background was integral to the reduction in criminal charges: "Were this crime committed by the defendant as someone who was born and raised in America, or born elsewhere but primarily raised in America, even in the Chinese American community, the Court would have been constrained to find the defendant guilty of manslaughter in the first degree" (cited in Volpp 1994, 73). Chen was convicted of second-degree manslaughter and received a sentence of 5 years probation with no jail time.

Such "cultural defense" cases can have the effect of denying equal protection of the laws to women, but do they really illustrate tensions between multiculturalism and gender equality? Would not most defenders of multiculturalism reject the use of the "cultural defense"? Indeed, most theorists of multiculturalism have not explicitly argued for the "cultural defense," and many "cultural defense" cases involve immigrants, who are not granted a substantial set of accommodations by some defenders of multiculturalism



because they are seen to have chosen to relinquish their cultural ties by immigrating (Kymlicka 1995, 30–31, 113–15; Spinner-Halev 2001, 87). But simply rejecting the “cultural defense” would deny immigrant defendants equal access to existing criminal defenses. Many “cultural defenses,” like the cases considered above, are not claims for complete exoneration or exemption from criminal laws; rather, they are requests for the consideration of cultural evidence in the application of existing criminal defenses, such as “mistake of fact” and “provocation.” What is being asked here is not special or differential treatment, but rather, similar treatment: minority defendants, like mainstream defendants, want to be judged in light of considerations about what, for example, it would be reasonable to be provoked by or what it would be reasonable to take as constituting consent. Because such factors depend on and may vary across cultural contexts, cultural evidence is needed to raise a “mistake of fact” or “provocation” defense. As Alison Dundes Renteln (2004, 32) has argued in defending the “cultural defense,” the “reasonable man” standard assumes the persona of the dominant culture and is thus “grossly unfair because it means that the provocation defense, which is supposed to be available to all is a defense only for those who belong to the dominant culture.” Simply jettisoning cultural evidence from the courtroom would deny minority defendants equal access to existing criminal defenses and may, therefore, jeopardize their rights to due process and equality before the law.<sup>5</sup> So, although most defenders of multiculturalism have not explicitly defended the “cultural defense,” a general case for multiculturalism that is based on the idea of equal treatment for cultural minorities provides a rationale for the “cultural defense.”

Feminist critics, including Okin (1998, 1999) and legal scholar Dorianne Lambelet Coleman (1996), have discussed these two “cultural defense” cases, among others, in order to illustrate the conflict between multiculturalism and gender equality and to argue against the former. Although feminists have rightly criticized the use of the “cultural defense” that leads to differential punishment for immigrant and American defendants, they have neglected the ways in which mainstream gender norms enable the accommodation of sexist practices within minority cultural communities. It is sometimes patriarchal mainstream norms that shape the frameworks within which minority claims are evaluated and granted or denied, and in the case of the “cultural defense,” cultural arguments seem to be most successful when they resonate with such norms. For example, although the defense lawyers in both the “marriage by capture” and “wife-murder” cases emphasized cultural *differences* between immigrants and mainstream Americans, there is a striking *congruence* in the norms of majority and minority cultures regarding the realm of intimate relations between the sexes.

In these cases, minority practices have found support from mainstream norms expressed in legal doctrine and practice. Thus, the problem here is not only the accommodation of cultural claims in criminal law but also mainstream gender norms.

Consider the “cultural defense” case involving Chen in light of the provocation defense for intimate homicide in mainstream law. American men who kill their wives or girlfriends have recourse to a criminal defense that, if successful, provides reduced charges and punishment. They are not called “cultural” defenses, but they rely on deeply rooted cultural understandings about what constitutes reasonable behavior between intimate partners. The “heat of passion” defense in traditional common law, or in Model Penal Code (MPC) terms the “extreme emotional disturbance” defense, is a partial excuse that mitigates murder to voluntary manslaughter. In trying to develop an objective standard of “provocation,” common law jurisdictions constructed specific common law categories of “adequate provocation,” including aggravated assault or battery, commission of a serious crime against a close relative of the defendant, and the observation of a spouse committing adultery (Dressler 1995, 491). What distinguishes adultery from other traditional common law categories of provoking events is that it does not involve an actual or threatened physical assault.

Today the great majority of American jurisdictions follow more open-ended provocation tests. This has led to two important related changes in provocation law, both favoring defendants: judges are much more likely to give a manslaughter instruction to juries, and juries now consider a wider variety of provoking conduct. About half of American jurisdictions still instruct in terms of “heat of passion,” whereas almost 20 states have adopted the MPC’s standard of “extreme emotional disturbance,” promulgated in 1962 (Lee 2003, 33, 285). In her study of the development and use of the “extreme emotional disturbance” standard, Victoria Nourse finds that the MPC reforms both broadened the range of relationships that might give rise to provocation claims (not just husband–wife but also boyfriend–girlfriend) and the types of conduct that might be classified as infidelity (not only having sexual relations with another but also trying to leave the relationship). Nourse (1997, 1342) notes that many states that base their test on the MPC have amended it substantially such that the great bulk of jurisdictions fall somewhere in between the traditional common law approach and the MPC approach.

The provocation defense, which can be traced back to the seventeenth-century conception of honor, appears to be gendered in both its formulation and its impact (Horder 1992). The substance of the legal rules concerning provocation that juries are asked to apply affects the behavior of all the decision makers involved in a homicide case. The open-endedness of these rules influence judges in deciding whether to give a provocation instruction and prosecutors in deciding what to charge and what kind of plea bargain to accept. There is an ongoing debate about whether current formulations of the provocation doctrine rely on what

<sup>5</sup> Legal scholar Leti Volpp also argues against abolition of the “cultural defense.” For her discussion of how cultural evidence might be considered in a way that is more critical than its current use, see Volpp 1996, 1611–13.

has been viewed as a masculine model of a sudden and temporary loss of self-control: a one-time-only encounter between two men of roughly equal size and strength (e.g., Crocker 1985; Maguigan 1991; Nourse 1997). In jurisdictions where provocation is formulated in this way, it may be less available to women subjected to long-term physical or sexual abuse who may act against their abuser sometime after his last assault or while he is sleeping. The question of explicit formulation of doctrine aside, what is harder to refute is the starkly disparate impact of the provocation defense along gender lines, which largely benefits men at the expense of women. Approximately 9% of homicides in the United States are committed by intimates, which the Department of Justice (DOJ) defines as current or former spouses, boyfriends, or girlfriends. A 1998 DOJ study found that it is increasingly the female rather than the male who is the victim in intimate homicides (Greenfeld et al. 1998, 1, 33). Men are arrested for more than 90% of all homicides, and almost three fourths of intimate homicide victims are female (Greenfeld et al., 5; Kaplan, Weisberg, and Binder 2000, 388–89; Rennison 2003, 1). The DOJ data tell us only about the incidence of intimate homicides, not about how many of these defendants claimed voluntary manslaughter or whether such claims were successful. But it also appears that men successfully utilize the provocation defense more often than do women. As the authors of one criminal law casebook put it, “Indeed, it is hard to find cases where a woman has her charge of punishment mitigated on provocation grounds when she has killed her husband or her husband’s lover” (Kaplan, Weisberg, and Binder 1996, 428; see also Kadish and Schulhofer 1995, 413).

In her study of 15 years of “heat of passion” and “extreme emotional disturbance” cases, Nourse finds that courts have extended the provocation doctrine to include not only a wife’s adultery but also the “infidelity of a fiancée who danced with another, of a girlfriend who decided to date someone else, and of the divorcee found pursuing a new relationship months after the final decree” (Nourse 1997, 1333, discussing *Dixon v. State*, 597 S.W. 2d 77 [Ark. 1980]; *Rodebaugh v. State*, No. 436, 1990 WL 254365 [Del. 1990]; *State v. Wood*, 545 A.2d 1026 [Conn. 1988]). Juries have returned manslaughter verdicts in cases where the defendant kills his wife and claims “passion” because the victim left him, moved the furniture out, planned a divorce, or sought a protective order (*State v. Little*, 462 A.2d 117 [N.H. 1983]; *State v. Reams*, 616 P.2d 498 [Or. 1980]; *Perry v. Commonwealth*, 839 S.W.2d 268 [KY 1992]; *Matthews v. Commonwealth*, 709 S.W.2d 414 [KY 1985]). According to Nourse (1343), “one is as likely, if not more likely, to find a relationship that has ended, was ending, or in which the victim sought to leave, as one is to find an affair or sexual infidelity alone.” In many states, current provocation law treats defendants who kill their intimate partners for trying to leave a relationship more favorably than in the past.

Today, U.S. law no longer views wives as the property of their husbands, and current formulations of the

provocation doctrine have jettisoned the language of honor in favor of “passion” or “extreme emotional disturbance.” But the provocation defense continues to operate in a way that reinforces the possessive norms rooted in the code of male honor: a woman’s infidelity, which in some jurisdictions includes her attempts to leave a relationship, betrays a loyalty expected of her. American courts have deemed such betrayal to be worthy of compassion and accommodation by the law. This was precisely the logic at the heart of the *Chen* case. Although the defense stressed the cultural differences in the ways that an American and Chinese man might respond to adultery, there is an underlying intercultural congruence in the gender norms at work: a man’s violent retaliation against his female partner’s infidelity is a reasonable response, whether committed out of honor, passion, or emotional disturbance.

Consider also the “wife-capture” case involving Moua in light of existing defenses in rape law. Not so long ago in the United States, unless there was obvious evidence of coercion, a woman charging rape had to convince the court that she had resisted the defendant’s advances “to the utmost.” In the absence of such resistance, the defendant could claim that he had made a “reasonable mistake” as to her consent. Many states have rewritten their laws minimizing the resistance requirement: rape laws no longer require that women resist “to the utmost”; “reasonable” resistance is supposed to be sufficient. Yet, out of a concern that defendants would have fewer clues as to nonconsent after the minimization of the resistance requirement, courts have been more willing than they have in the past to admit a “mistake of fact” defense (Kadish and Schulhofer 1995, 327). Rape traditionally has involved two elements: force on the part of the perpetrator and lack of consent on the part of the victim. In most states, a defendant charged with rape can raise a “mistake of fact” defense, which allows him to claim that his belief as to the other party’s consent was honest and reasonable. Most rape statutes still use some combination of “force,” “threats,” and “consent” to define the threshold of liability—the line between criminal sex and seduction (Estrich 1986, 1184). In giving meaning to those terms at the threshold of liability, the law of rape continues to draw upon very powerful mainstream norms of male aggressiveness and female passivity.

In Moua’s case, the defense lawyer did not explicitly invoke the mistake of fact defense, but in response to the district attorney’s assertion that he had “never heard of any other cultures getting a break because they thought [rape or kidnapping] was okay,” Moua’s lawyer replied that “in the California culture” defendants have been given some “credit” by the courts and cited *People v. Mayberry*, 542 P.2d 1337 [Cal. 1975] (cited in *People v. Moua* [1985], 7). This case involved a man who approached a woman at a local store, propositioned her for sex, and demanded that she go back with him to his home. When she refused, he struck her. The store personnel and other customers did not see them. Out of fear, she did not resist his demand to

accompany him back to his home, and during the sexual assault in his home, she did not resist. The California Supreme Court reversed Mayberry's kidnapping and rape conviction on the grounds that in the absence of resistance, it was reasonable for him to believe she had consented to sex. This case is not an exception; in nearly all states, intimidation short of physical threats, including psychological pressure used by people in positions of authority over their subordinates, is treated as if it were mere persuasion. For example, a high school principal who told a student that he would not allow her to graduate if she did not have sex with him was not held in violation of law (*State v. Thompson*, 792P2d 1103 [Mont. 1990]). In such cases involving intimidation short of physical threats, courts usually say the victim consented. The vast majority of states have no law requiring courts to accept verbal refusal at face value (Schulhofer 1998, 11). This is because the old idea that women who say "no" to sexual advances don't really mean no is still widely accepted in the majority culture.<sup>6</sup> Some Hmong men may engage in cultural practices that subordinate Hmong women, but there are similarly powerful norms of male aggressiveness and female passivity at work in mainstream legal doctrine and practice, and the latter have offered support for the former.

In sum, the problem raised by the "cultural defense" cases above has not only to do with minority practices but also with mainstream norms that offer support for those practices. One might argue for eliminating the extra reduction of charges and punishment that derives from the use of cultural evidence and leave it at that. Both Coleman and Okin argue for eliminating the "cultural defense" on the grounds that it withholds equal protection of the laws to victims from the cultures being accommodated. But even if the extra reduction in punishment based on cultural evidence were eliminated, the majority culture's own gendered understandings of agency and responsibility would remain, where men are seen as reasonably provoked to kill by the sexual behavior of their partners and where women who say "no" don't really mean no. As Anne Phillips (2003) has argued in examining the "cultural defence" in English courts, the larger problem with the use of cultural evidence is that they have proven to be most effective when they resonate with mainstream norms. So long as gender-biased norms pervade legal doctrine and practice, minority defendants can continue to find support for patriarchal practices within mainstream law. Adequately capturing and responding to the problems raised by the cultural defense, then, requires being attentive to the ways in which majority norms shape the practice of accommodation.

<sup>6</sup> In contrast to most states, the state of Massachusetts does not allow "mistake of fact" defenses in rape cases. Instead, it has adopted a strict liability standard under which the legal definition of rape could include cases where a person says no but does not physically resist, or where she submits in response to lies or threats. See *Commonwealth v. Ascolillo* (1989), 405 Mass. 456, 541 N.E.2d 570 (1989).

## Minority → Majority: Boomerang Effect

Majority norms have influenced the gender norms of minority cultures, but influence can also run in the other direction. The legal accommodation of sexist practices within minority cultures, as in the "cultural defense" cases, may boomerang back to threaten struggles toward gender equality within the wider society. This is the one interactive dynamic that Okin (1999) and Coleman (1996) have stressed in their critique of multiculturalism. "Cultural defenses" appear to have had mixed success in criminal cases across federal and state jurisdictions, and none, to my knowledge, has been cited in cases involving defendants of the dominant culture. But when courts rule in ways that tolerate sexist practices among immigrants, as some courts have, those cases may well validate sexist norms in the majority culture. Such "cultural" cases become potential precedents, and this fact alone means that mainstream law has been reshaped. In making the case for a jury instruction of provocation, a mainstream defendant could point to "cultural" cases and argue that if immigrants can have access to the provocation defense, then he should, too. If he were denied access to the provocation defense, he could again point to "cultural" cases and argue that he is being denied protections that immigrant defendants enjoy. It is also important to be attentive to potential boomerang effects for the following reasons.

First, although several recent federal cases suggest that the boomerang effect is small, judges have left the door open to the use of culture in the courts. For instance, in a 2001 case involving a Mexican woman convicted of a drug charge, Judge Richard Posner reversed a reduction in punishment granted by a sentencing judge on the basis of the defendant's "cultural heritage." Maira Bernice Guzman sought a reduced sentence for a drug charge for which she and her boyfriend had been convicted; she sought a reduced sentence on the grounds that "Mexican cultural norms dictated submission to her boyfriend's will" (*U.S. v. Guzman*, 236 F. 3d 830 [2001], 830–31; see also *U.S. v. Contreras*, 180 F.3d 1204 [1999]; *U.S. v. Natal-Rivera*, 879 F.2d 391 [1989]). Judge Posner argued that to mitigate punishment on the basis of cultural evidence would be an "abuse of discretion" because the U.S. Sentencing Guidelines prohibits consideration of race, sex, national origin, creed, religion, and socioeconomic status in determining sentences. Although "culture" or "ethnicity" is not specified in the guideline, he suggests that the drafters thought that the stated exclusions encompassed culture and ethnicity. Giving judges leeway to consider "cultural heritage" in sentencing decisions, Judge Posner argues, "would inject enormous subjectivity and variance into a sentencing scheme designed to achieve reasonable objectivity and uniformity." But he leaves the door open to the use of cultural evidence in future sentencing cases, arguing that prohibition would "exclude all possibility of consideration of cultural factors in cases that we cannot yet foresee" (833–34).

Second, the “cultural defense” has not been limited to cases where the parties involved are from the same culture. For instance, in *Gonzales v. State*, 689 S.W.2d 900 [Tex. 1985], the defendant was convicted of murder for fatally shooting his wife after a heated argument and sought a jury instruction that the situation be assessed from his own perspective, that of “a Hispanic farm worker who was living with a Caucasian woman on a low income.” The trial judge rejected the defendant’s proposed jury instruction, but this case is, nonetheless, troubling since judges exercise considerable discretion on whether and how cultural evidence gets considered, and also because the proposed jury instruction reflects the idea, increasingly made by minority defendants, that equal access to mainstream legal defenses requires consideration of cultural factors, including patriarchal traditions, in explaining a minority defendant’s state of mind. This is precisely what a second-generation Japanese-American man argued in the recent unpublished California state appellate case, *People v. Kobayashi*, No. B157685 [Cal. App. 2 Dist. 2003]. Kobayashi sought to overturn his conviction for murdering Sheila Ann Randle, an African American woman with whom he had had a relationship. On appeal, he argued that the jury should have been instructed “to evaluate the sufficiency of provocation from the standpoint of a reasonable person in terms of defendant’s position as a Japanese American” (10). The expert psychologist in the case linked the defendant’s state of mind with his cultural background: “[I]n Japanese culture, intense shame attaches to males who lack emotional control, who are unable to meet the expectations of others, and who violate their personal standards” (9). Kobayashi argued that “equal treatment of ethnic minority defendants requires that if certain provocative acts are sufficiently offensive in mainstream American culture to reduce murder to manslaughter, then certain acts that are equally provocative in appellant’s culture should be treated as equally mitigating” (11). The state appellate court upheld the conviction, sidestepping the question—whether there is an equal protection and due process right to a culturally specific evaluation of the element of provocation as it relates to the crime of manslaughter—on the grounds that it had not been raised during trial.

Third, while no “cultural defense” cases have been cited as precedents in cases involving defendants of the dominant culture, one published “cultural defense” case has been cited as a precedent in another “cultural defense” case, suggesting that “cultural” cases are not always a one-off matter and that boomerang effects can occur across minority groups. A federal appellate court held that cultural evidence may be admitted where it is relevant to the defendant’s culpability for the crimes alleged (*Bains v. Cambra*, 204 F. 3d 964 [2000]). In this case, a Sikh man Bains was convicted as a co-conspirator in the murder of his sister’s ex-husband Shergill. Cultural evidence was offered by the prosecution to make the case that Bains was motivated to kill Shergill in part because of his Sikh religion. Several witnesses testified that Sikh families “feel very

strongly that a husband must comply with his half of the marriage contract, especially since if a husband leaves his wife, his wife is considered to be ‘damaged goods’ and an ‘unmarketable commodity,’ thereby causing the families of both spouses great hardship” (970). The *Bains* court permitted the use of cultural evidence in order to elucidate a possible motive for Bains to have Shergill killed. This case was then cited in a case involving an Indian immigrant Hundal, who had been convicted of rape and spousal abuse (*People v. Hundal*, No. F037541 [Cal. App. 5 Dist. 2002]). His wife had used cultural evidence to explain why she had been willing to agree to an arranged marriage and to stay with him despite a history of physical and sexual abuse. Hundal sought to overturn his conviction on the grounds that the prosecutor’s stereotypical characterization of “Indian culture”—that “men control women” and “have a higher status than women” in “Indian culture”—had denied him a fair trial. The Court upheld the conviction on the grounds that the prosecutor’s improper questioning had not affected the jury’s verdict and that the use of cultural evidence had been entirely proper because the prosecutor’s questions about “whether appellant himself thought of the victim as an item of property were relevant to the charges at hand” (6).

In *Bains*, a man sought the admittance of cultural evidence to overturn his conviction for avenging what he understood to be his sister’s dishonored status; in *Hundal*, a woman invoked cultural evidence to explain why she did not leave an abusive relationship in bringing a rape charge against her husband. But in both cases, courts permitted juries to consider patriarchal traditions to explain people’s behavior. In both cases, juries chose to convict. But prosecutors and judges exercise considerable discretion in whether and how culture gets used in the courtroom, and federal and state jurisdictions have increasingly permitted juries to consider “cultural defenses,” including evidence of patriarchal cultural traditions, to explain defendants’ behavior. In some locales, juries have allowed such defenses to serve as partial excuses for patriarchal behavior among immigrants, and mainstream defendants can and may well point to these “cultural” cases in raising their own criminal defense claims. Given the increasingly diverse immigrant presence in the United States and the increasing use of cultural defenses, it is important to be attentive to boomerang effects.

### Minority → Majority: Diversionsary Effect

Even in cases where minority claims for accommodation are denied, we need to be attentive to interactive dynamics between majority and minority cultures. In some cases, rejection and vilification of minority claims for accommodation can serve to divert attention from the majority culture’s own practices. This diversionsary effect can be seen in the case of Mormon polygamy in nineteenth-century America, as well as in contemporary debates over gay marriage and female circumcision. In the case of Mormon polygamy, even as American reformers and government officials

resisted the ideas of feminism, they deployed the language of feminism in the service of its efforts to dismantle polygamy. Such rhetoric not only provided them with a ready justification for intervention into Mormon communities but also helped shield mainstream gender practices—Christian-model monogamy and the traditional gender roles associated with it—from criticism.<sup>7</sup>

The U.S. government's efforts to dismantle Mormon polygamy spanned from 1862 to 1890. It pursued the campaign against polygamy with, as one legal historian puts it, "a zeal and concentration" that have been "unequalled in the annals of federal law enforcement" (Linford 1964, 312, 585). In 1862, Congress criminalized bigamy in the territories, and in 1879, the Supreme Court upheld this law (Morrill Anti-Bigamy Act, 12 Stat. 501 [1862]; *Reynolds v. United States*, 98 U.S. 145 [1879]). Congress followed up in 1887 with the Edmunds-Tucker Act, which repealed the incorporation of the Mormon Church, directed the U.S. attorney general to escheat its property holdings over \$50,000, and disenfranchised Mormon women (Edmunds-Tucker Act, 24 Stat. 635 [1887]). The Mormons resisted and continued to practice polygamy, but in 1889, the Supreme Court upheld Congress's power to dissolve and expropriate the church's property against the church's claim that it was a protected religious body (*Late Corporation of the Church of Jesus Christ of Latter-day Saints v. U.S.*, 136 U.S. 1 [1889]). Finally, in 1890, Mormon President Wilson Woodruff issued a manifesto accepting the federal prohibition of polygamy and encouraged fellow Mormons to refrain from contracting any further polygamous marriages.

Nineteenth-century antipolygamy arguments given by politicians, judges, and activists focused on what they believed was a deeply patriarchal practice. Writing for the majority in *Reynolds v. U.S.* (1879), Chief Justice Morrison Waite expressed concern for the "pure-minded women" who were the "the innocent victims of this delusion," and argued for proscribing polygamy on the grounds that it "leads to the patriarchal principle . . . which, when applied to large communities, fetters the people in stationary despotism, while that principle cannot long exist in connection with monogamy" (*Reynolds v. United States*, 98 U.S. 145 [1879], 165–68). The association between monogamy and freedom, on the one hand, and polygamy and despotism, on the other, can be traced back to Montesquieu's ([1748] 1989, 270, 316) idea that "do-

mestic government" shaped "political government." The leading political and legal philosophers of early and antebellum America, from James Wilson to Chancellor James Kent, Supreme Court Justice Joseph Story, and Francis Lieber, all contrasted monogamy with polygamy in order to illustrate the superiority of Christian morality over "oriental despotism" (Cott 2000, 22–23). To be sure, as echoed by the *Reynolds* Court, antipolygamy activists and politicians were motivated by a concern to dismantle what they believed was a deeply patriarchal institution, but Mormon sexual practices were not their sole concern.

The association of female enfranchisement and liberal divorce laws with Mormon sexual practices fueled fears that all three were part of a plot to undermine the traditional American family and Christian civilization itself (Gordon 2002, 52–54; Smith 1997, 388). The Mormon-controlled territorial legislature in Utah had unanimously approved the enfranchisement of women in 1870, making the women of Utah among the first women to vote in America (Foster 1981, 214; Keyssar 2000, 54; Smith, 106, 110). By the time they were disenfranchised by Congress, women in Utah had had the right to vote for 17 years. Opponents of polygamy and opponents of women's suffrage found common cause when it came to Mormonism. Francis Lieber, whose antipolygamy arguments Chief Justice Waite cited in writing the *Reynolds* opinion, opposed women's suffrage on the grounds that women's suffrage, like polygamy, would undermine the American family ([1839] 1875, 124–125).

In addition to women's suffrage, the ease of divorce in Utah added fuel to antipolygamists' condemnations that Mormon polygamy would wreck the American family. To them, polygamy and divorce were linked; both treated marriage as temporary or vulnerable to whim. Antidivorce activists referred to divorce as "the polygamic principle" or "polygamy on the instalment [sic] plan" (cited in Gordon 1996, 836). Anxiety over rising divorce rates added momentum to the family law reform movement, whose supporters overlapped with supporters of the antipolygamy movement. Antidivorce and antipolygamy activists joined forces in calling for a uniform national marriage law. In 1886, Senator George Edmunds, Congress's leading antipolygamy spokesman, also attempted to get a bill through Congress which would authorize the government to collect divorce statistics as a first step toward restricting divorce (Iversen 1997, 106–107). According to these reformers who sought both to eliminate polygamy and restrict divorce, monogamous marriage and women's traditional place within it was the basis of civilized society and democratic government.

Women's rights activists joined forces with Mormon women defending polygamy, making for a doubly threatening combination in the eyes of defenders of Christian monogamy. Shortly after they were enfranchised, Mormon women established contact with leading women's rights activists, and as early as 1872, Mormon women held office in the National Woman Suffrage Association (NWSA), the suffrage

<sup>7</sup> The diversionary effect of focusing on "other" men's treatment of women has been used to justify intervention by states not only into minority cultures within one state but also intervention across states. In examining the conduct and rhetoric of the British colonial establishment toward Islamic societies, Leila Ahmed shows how British officials appropriated the language of feminism in the service of colonialism. The result was the fusion of the issues of women's oppression and the cultures of "other" men such that improving the status of women was thought to entail abandoning native customs. She also argues that this focus on "other" men's treatment of women helped Western colonial governors to oppose feminism within their own societies in favor of more traditional gender roles (Ahmed 1992).

organization led by Susan B. Anthony and Elizabeth Cady Stanton. Stanton and Anthony were invited to speak in Salt Lake City. In her lecture from the pulpit of the Mormon Tabernacle in 1871, Stanton attacked patriarchal power and the subordination of women by organized religion and argued that there was just as good reason for polyandry as there was for polygyny (Iversen 1997, 25–26). Accompanied by NWSA members, Mormon women delivered a memorial to the House Judiciary Committee on behalf of all Mormon women, defending their practice of polygamy in the name of women's rights and asking Congress to repeal the Morrill Act of 1862. They maintained that Mormon women were contented wives and mothers and that the effect of enforcing antipolygamy legislation would make 50 thousand women outcasts and their children illegitimate (Iversen, 29–30). The alliance between NWSA and Mormon women was possible in part because the NWSA members rejected the claim that it was the *form* of marriage that subordinated women.

By the time the issue of polygamy arose on the national political stage, nineteenth-century women's rights activists had already been unsettling prevailing gender norms. By the mid-nineteenth century, women's rights activists had begun criticizing women's subordination within all forms of marriage. Quaker abolitionist Angelina Grimké's examination of slavery led her to conclude that all married women endured many of the same legal and social disabilities as slaves. The Declaration of Sentiments drafted by Stanton for the Seneca Falls Convention of 1848 made legal demands for women's rights, attacking the doctrine of coverture. Starting in 1839, states passed legislation challenging coverture, including Married Women's Property Acts, which permitted women to inherit, own, and exchange property independently of their husbands. By 1865, twenty-nine states had passed some form of married women's property law (Basch 1982, 28; Kerber 1998, 38–39).

It is in this larger context of mainstream gender practices that Stanton and Anthony viewed the controversy over Mormon polygamy. Stanton herself ([1871] 1970, 70) distinguished among three kinds of polygamy: Mormon polygamy, bigamy based on fraud, and polygamy involving one man, one wife, and many mistresses "everywhere practiced in the United States." Rather than condemn Mormon polygamy and defend Christian monogamy, Stanton criticized all contracts of marriage as oppressive for women: "In entering this contract, the man gives up nothing that he before possessed—he is a man still; while the legal existence of the woman is suspended during marriage, and henceforth she is known but in and through the husband." She sought to improve women's status by arguing for greater equality within marriage and greater freedom to divorce (Stanton, Anthony, and Gage [1881–1922] 1970, 738–40; Clark 1990, 34–38). Similarly, Anthony urged suffragists to avoid "shouts of puritanic horror" against polygamy and offer a "simple, loving, sisterly clasp of hands" in order to help abolish "the whole system of woman's subjection to man in both polygamy

and monogamy" (quoted in Iversen 1997, 35). The NWSA never publicly condemned polygamy in part because they rejected the claim that polygamy was inherently patriarchal and monogamy inherently egalitarian, and also because Mormons had enfranchised women and provided women with greater freedom to divorce. When the federal government moved to disenfranchise the women of Utah on the grounds that they were brainwashed by their husbands, NWSA activists argued against the use of "federal power to disenfranchise the women of Utah, who have had a more just and liberal spirit shown them by Mormon men than Gentile women in the States have yet perceived in their rulers" (Stanton, Anthony, and Gage, 128).

Antipolygamists who sought to defend Christian monogamy in the face of attacks by women's rights activists found a convenient diversion in Mormon polygamy. As legal historian Sarah Barringer Gordon (2002, 54) puts it, "The popular appeal of antipolygamy gave legislators a convenient out—here was a form of marriage that *truly* replicated 'slavery' for white women. By enacting laws to prohibit the 'enslavement of women in Utah,' congressmen could deflect attention from domestic relations in their own states and direct it towards a rebellious territory. In this sense, Utah became a handy foil." To be sure, public officials and citizens attacked polygamy because they believed it to be a deeply patriarchal practice, but their focus on polygamy had the effect of shielding patriarchal aspects of monogamous marriage from the criticism of women's rights activists. Diverting attention from monogamy may not have been antipolygamists' primary intent, but the focus on polygamy served the cause of those who defended Christian-model monogamy and the traditional gender roles associated with it.<sup>8</sup>

Polygamy is no longer mandated by the Mormon Church, but the diversionary effect is not a relic of the past. Although official church doctrine condemned polygamy starting in 1890, some Mormons continue to practice it. Today about 20,000 to 50,000 Mormon fundamentalists live in polygamous families, and the government has largely taken a "don't ask, don't tell" policy toward them (Altman and Ginat 1996, 51, 54). Although gender hierarchies may indeed pervade such communities, like the nineteenth-century women's rights activists, we should be wary of concluding that polygamy is necessarily patriarchal and monogamy necessarily egalitarian. The charge that polygamy is oppressive to women is contingent and thus needs to be investigated by looking at individual relationships and their context, just as we ought to do with monogamous relationships (see Emens 2004).

The diversionary effect can also be seen beyond the case of polygamy. Consider the contemporary debate

<sup>8</sup> The focus on polygamy was not only a handy foil against NWSA and other women's rights activists' critiques of monogamy, but also a diversion from the federal government's attack on what was probably its bigger concern: the great political power of the Mormon Church in Utah, which, in President Rutherford B. Hayes's words, "bore no semblance to republican government." See Rosenblum 1997, 75–6.



on same-sex marriage in the United States. Both proponents and opponents of same-sex marriage seem to embrace the long-standing view of family as based on romantic monogamous sexual affiliation. But this embrace can serve to shield the institution of monogamous marriage from more radical challenges—for instance, that marriage is designed to privilege those inside and discipline those outside it: single people with or without dependents, unwed parents, divorcees, prostitutes, or those who do not desire legal recognition of their intimate relationships (Warner 2002, 264–65). The movement for same-sex marriage also leaves the privilege of spousal status intact, masking important issues concerning dependency and care-giving for those who are in need, including children, many elderly, the ill, and the disabled (Fineman 2004, 48). The denial of privileges to other intimate and care-giving relationships largely leaves women, who do much of the care-giving not only within but also outside marriage relations, to bear the burdens of caring for dependents. With its focus on the good of marriage and opening up access to it, the contemporary American debate on same-sex marriage, as with the nineteenth-century American debate on polygamy, may serve to deflect attention from more radical challenges to marriage and the traditional gender roles long associated with it.

Similarly, condemnation of various practices among immigrants may help divert attention from majority practices. Focusing on cases of domestic violence or rape in immigrant communities can serve to reinforce the false dichotomy of oppressive minority cultures and egalitarian majority cultures, deflecting attention from the reality of domestic violence and other patriarchal practices within majority cultures (An-Na'im 1999, Volpp 2001). Western feminists' criticism of female circumcision among immigrant communities can serve to divert attention from the variety of cosmetic surgeries and bodily alterations, such as breast enlargement, facelifts, and labiaplasty, as well as male circumcision, performed in Western societies (Carens and Williams 1998, 481; Navarro 2004). There may well be a good case that certain forms of female circumcision should be treated differently from other forms of bodily alterations commonly practiced in the West. The point here is that majority norms and practices should be analyzed alongside minority norms and practices, lest we overlook the diversionary effects that can reinforce gender hierarchies across cultures and also fuel discourses of cultural and racial superiority within the dominant culture.

## CONCLUSION

At least two important implications follow from the interactive dynamics illustrated by the cases discussed above for the theory and practice of multiculturalism. The first has to do with how conflicts of culture are understood, which has in part to do with how culture has been conceptualized in the debate on multiculturalism. The formulation of cultures as well-integrated, well-bounded, and self-generated entities suggests an

*internal view* of the dilemmas of gender and culture. For instance, in diagnosing the problems raised by multiculturalism for feminism, Okin, like Kymlicka, seems to view cultures as unified wholes, even while recognizing gender as a social cleavage. For her, cultures are, on the whole, deeply patriarchal structures. As she puts it (1999, 14, 17), "Many of the world's traditions and cultures, including those practiced within formerly conquered or colonized nation-states . . . are quite distinctly patriarchal . . . Most cultures are patriarchal, then, and many (though not all) of the cultural minorities that claim group rights are more patriarchal than the surrounding cultures." But cultures are less monolithically patriarchal than Okin suggests, and minority cultures are not straightforwardly "more patriarchal" than Western majority cultures (see Honig 1999, Narayan 1997, Volpp 2001). A view of cultures as distinct wholes encourages an internal view of the gendered conflicts of culture, overlooking sources of minority women's subordination that do not stem from within their own cultural communities but from structural forces beyond their communities. As I have tried to show, this internal view fails to capture a range of interactive dynamics that shape cultural conflicts—in particular, the state's role in encouraging the accommodation of minority practices that are congruent with the dominant culture's own norms and practices. As Okin (1989) herself demonstrates in *Justice, Gender, and the Family*, the laws and norms of the dominant culture continue to reinforce gender injustice in certain respects. In some cases, it has been the congruence of patriarchal norms and practices between majority and minority cultures that has supported accommodation. If we take a social constructivist view of cultures as not only internally contested but also interactive and mutually constitutive, then the problem for feminists is not so much "multiculturalism versus gender equality" as "patriarchy versus gender equality." Formulating the problem in this way allows for consideration of the role that mainstream norms and institutions play in shaping minority practices. It also allows for examination of non-cultural factors, such as structures of class and race, that impede struggles toward gender equality.

The second implication has to do with how to formulate responses to the dilemmas associated with culture. If we conceive of cultural conflicts as "multiculturalism versus gender equality," the only remedy appears to be either siding with multiculturalism at the expense of gender equality or siding with gender equality at the expense of multiculturalism. This either/or strategy falls short. On the one hand, as feminist theorists have stressed, simply saying yes to cultural accommodation altogether fails to protect vulnerable members within minority cultural communities. But simply saying no to accommodation and demanding assimilation (or saying no until minority cultures liberalize up to the level of majority cultures) also falls short because the majority culture is in certain respects not less patriarchal than minority cultures and because rejecting accommodation altogether may fail to accord equal respect to members of minority cultures.

Adopting an interactive view of cultural conflicts, as I have argued for here, suggests the need to develop more context-sensitive approaches to evaluating the claims of minority cultures. Rather than relying on dichotomous generalizations about oppressive minority cultures and egalitarian Western cultures, evaluations of minority claims should be based on examination of particular practices in particular contexts with an eye toward actual and potential interconnections between minority and majority practices. Taking this interactive view does not mean accepting that majority cultures are always implicated in the maintenance of hierarchies within minority cultures. What it requires, in contrast to the internal view, is vigilance toward the ways that mainstream institutions and norms have been and could be so implicated.

If one defends special protections for minority cultural groups, on grounds of equality or autonomy, evaluation of minority claims ought to proceed with the following sorts of contextual considerations in mind: (1) What is the origin and history of the practice in question, and what is its importance to members of the cultural group? What role, if any, has the state played in its development? In the Santa Clara Pueblo case, the tribe's interest in maintaining itself as a "culturally and political distinct entity" would have to be weighed against the interests of individuals seeking to overturn the gender-biased membership rule. But, rather than taking tribal claims for the distinctness and importance of a tribal tradition at face value, consideration of the claim should include investigation of the tradition's origins and the state's role in its creation and maintenance. (2) How do mainstream laws and norms influence which minority claims get accommodated? As in the case of the "cultural defense," culture may well operate within evaluative frameworks already defined by the majority culture's own norms, which in some cases are patriarchal. This suggests the need to scrutinize mainstream legal doctrines and practices, alongside minority practices, in considering whether and how cultural evidence should be used. Arguments for limiting the use of "cultural defenses" that reinforce gender-biased norms and practices, as in the Chen and Moua cases, must be tied to critical reexamination of mainstream legal defenses implicated in these "cultural" cases. (3) What effects will granting or denying minority claims have on members of the minority groups being accommodated, as well as on members of other minority groups and the dominant culture? Will criticism or accommodation of particular minority practices reinforce gender inequality or play into cultural and racial stereotypes?

Such context-sensitive considerations would infuse evaluations of minority claims with greater cross-cultural humility. This does not mean that democratic citizens should look the other way in the face of oppressive practices. Rather, it means acknowledging the ways in which struggles toward gender equality within Western majority cultures are incomplete and ongoing, and being vigilant of the interconnections between such struggles and those within minority communities.

## REFERENCES

- Ahmed, Leila. 1992. *Women and Gender in Islam: Historical Roots of a Modern Debate*. New Haven, CT: Yale University Press.
- Altman, Irwin, and Joseph Ginat. 1996. *Polygamous Families in Contemporary Society*. Cambridge: Cambridge University Press.
- An-Na'im, Abdullahi. 1999. "Promises We Should All Keep in Common Cause." In *Is Multiculturalism Bad for Women?* eds. Joshua Cohen, Matthew Howard, and Martha C. Nussbaum. Princeton, NJ: Princeton University Press, 59–64.
- Anthony, Susan B. [July 20, 1871] 1992. "The Liberal Party at Salt Lake." *The Revolution* 8. In *The Papers of Elizabeth Cady Stanton and Susan B. Anthony* [1831–1906], eds. Patricia G. Holland and Ann D. Gordon. Wilmington, DE: Scholarly Resources. Microform.
- Appiah, Kwame Anthony. 2004. *The Ethics of Identity*. Princeton, NJ: Princeton University Press.
- Bains v. Cambra, 204 F.3d 964 [2000].
- Basch, Norma. 1982. *In the Eyes of the Law: Women, Marriage, and Property in Nineteenth-Century New York*. Cornell, NY: Cornell University Press.
- Benhabib, Seyla. 2002. *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton, NJ: Princeton University Press.
- Bredbenner, Candace. 1998. *A Nationality of Her Own: Women, Marriage, and the Law of Citizenship*. Berkeley: University of California Press.
- Carens, Joseph H., and Melissa S. Williams. 1998. "Islam, Immigration, and Group Recognition." *Citizenship Studies* 2 (3): 475–500.
- Clark, Elizabeth B. 1990. "Matrimonial Bonds: Slavery and Divorce in Nineteenth-Century America." *Law and History Review* 8 (1): 25–52.
- Coleman, Doriane Lambelet. 1996. "Individualizing Justice through Multiculturalism: The Liberals' Dilemma." *Columbia Law Review* 96 (June): 1093–167.
- Commonwealth v. Ascolillo* (1989), 405 Mass. 456, 541 N.E.2d 570 (1989).
- Cott, Nancy F. 1998. "Marriage and Women's Citizenship in the United States, 1830–1934." *American Historical Review* 103 (5): 1440–75.
- Cott, Nancy F. 2000. *Public Vows: A History of Marriage and the Nation*. Cambridge, MA: Harvard University Press.
- Crocker, Phyllis L. 1985. "The Meaning of Equality for Battered Women Who Kill Men in Self-Defense." *Harvard Women's Law Journal* 8: 121–53.
- Deveaux, Monique. 2000. "Conflicting Equalities? Cultural Group Rights and Sex Equality." *Political Studies* 48 (3): 522–39.
- Deveaux, Monique. 2003. "A Deliberative Approach to Conflicts of Culture." *Political Theory* 31 (6): 780–807.
- Dixon v. State*, 597 S.W.2d 77 [Ark. 1980].
- Dozier, Edward P. 1970. *The Pueblo Indians of North America*. New York: Holt, Rinehart and Winston.
- Dressler, Joshua. 1995. *Understanding Criminal Law*. 2nd ed. New York: Matthew Bender/Irwin.
- Emens, Elizabeth F. 2004. "Just Monogamy?" In *Just Marriage*, eds. Joshua Cohen and Deborah Chasman. Oxford: Oxford University Press, 75–80.
- Estrich, Susan. 1986. "Rape." *Yale Law Journal* 95: 1087–163.
- Fineman, Martha Albertson. 2004. "Why Marriage?" In *Just Marriage*, eds. Joshua Cohen and Deborah Chasman. Oxford: Oxford University Press, 46–51.
- Foster, Lawrence. 1981. *Religion and Sexuality: The American Communal Experiments of the Nineteenth Century*. Oxford: Oxford University Press.
- Gonzales v. State*, 689 S.W.2d 900 [Tex. 1985].
- Gordon, Sarah Barringer. 1996. "'The Liberty of Self-Degradation': Polygamy, Woman Suffrage, and Consent in Nineteenth-Century America." *Journal of American History* 83 (December): 815–47.
- Gordon, Sarah Barringer. 2002. *The Mormon Question: Polygamy and Constitutional Conflict in Nineteenth-Century America*. Chapel Hill, NC: University of North Carolina Press.
- Green, Leslie. 1995. "Internal Minorities and Their Rights." In *The Rights of Minority Cultures*, ed. Will Kymlicka. Oxford: Oxford University Press, 257–72.

- Green, Rayna. 1980. "Native American Women." *Signs* 6 (2): 248–67.
- Greenfeld, Lawrence A., et al. 1998. *Violence by Intimates. Analysis of Data on Crimes by Current or Former Spouses, Boyfriends, and Girlfriends 1976–1996*. Washington, D.C.: U.S. Dept. of Justice, Office of Justice Programs.
- Hill, Willard Williams. 1982. *An Ethnography of Santa Clara Pueblo, New Mexico*. Albuquerque: University of New Mexico Press.
- Honig, Bonnie. 1999. "My Culture Made Me Do It." In *Is Multiculturalism Bad for Women?* eds Joshua Cohen, Matthew Howard, and Martha C. Nussbaum. Princeton, NJ: Princeton University Press, 35–40.
- Horder, Jeremy. 1992. *Provocation and Responsibility*. Oxford: Clarendon Press.
- Iversen, Joan Smyth. 1997. *The Antipolygamy Controversy in U.S. Women's Movements, 1880–1925: A Debate on the American Home*. New York: Garland Publishing.
- Jacobs, Margaret D. 1999. *Engendered Encounters. Feminism and Pueblo Cultures, 1879–1934*. Lincoln: University of Nebraska Press.
- Kadish, Sanford H., and Stephen J. Schulhofer. 1995. *Criminal Law and Its Processes Cases and Materials*. 6th ed. Boston: Little, Brown, and Co.
- Kaplan, John, Robert Weisberg, and Guyora Binder. 1996. *Criminal Law: Cases and Materials*, 3rd ed. Boston: Little, Brown, and Co.
- Kaplan, John, Robert Weisberg, and Guyora Binder. 2000. *Criminal Law: Cases and Materials*, 4th ed. Boston: Little, Brown, and Co.
- Kerber, Linda. 1998. *No Constitutional Right to Be Ladies. Women and the Obligations of Citizenship*. New York: Hill and Wang.
- Keyssar, Alexander. 2000. *The Right to Vote: The Contested History of Democracy in the United States*. New York: Basic Books.
- Kymlicka, Will. 1989. *Liberalism, Community, and Culture*. Oxford: Oxford University Press.
- Kymlicka, Will. 1995. *Multicultural Citizenship: A Liberal Theory of Minority Rights*. Oxford: Oxford University Press.
- Kymlicka, Will. 2001. *Politics in the Vernacular: Nationalism, Multiculturalism and Citizenship*. Oxford: Oxford University Press.
- Late Corporation of the Church of Jesus Christ of Latter-day Saints v. U.S.*, 136 U.S. 1 [1889].
- Lee, Cynthia. 2003. *Murder and the Reasonable Man: Passion and Fear in the Criminal Courtroom*. New York: New York University Press.
- Leonard v. Grant*, 5 F.11 [1880].
- Levy, Jacob. 1997. "Classifying Cultural Rights." In *NOMOS XXXIX: Ethnicity and Group Rights*, eds Ian Shapiro and Will Kymlicka. New York: New York University Press, 22–66.
- Lieber, Francis. [1839] 1875. *Manual of Political Ethics*. 2nd ed. rev. and ed Theodore D. Woolsey. Philadelphia: J. B. Lippincott & Co. Microfiche.
- Linford, Orma. 1964. "The Mormons and the Law: The Polygamy Cases, Part I." *Utah Law Review* 9 (Winter): 308–70.
- MacKinnon, Catherine A. 1987. "Whose Culture? A Case Note on *Martinez v. Santa Clara Pueblo*." In *Feminism Unmodified: Discourses on Life and Law*. Cambridge, MA: Harvard University Press, 63–9.
- Maguigan, Holly. 1991. "Battered Women and Self-Defense: Myths and Misconceptions in Current Reform Proposals." *University of Pennsylvania Law Review* 140 (December): 379–486.
- Margalit, Avishai, and Moshe Halbertal. 1994. "Liberalism and the Right to Culture." *Social Research* 61 (3): 491–510.
- Martinez v. Santa Clara Pueblo*, 402 F. Supp. 5 [1975].
- Martinez v. Santa Clara Pueblo*, 540 F.2d 1039 [1976].
- Matthews v. Commonwealth*, 709 S.W.2d 414 [KY 1985].
- Moapa Band of Paiute Indians v. U.S. Department of Interior*, 747 F.2d 563 [1984].
- Montesquieu. [1748] 1989. *Spirit of the Laws*. Trans. and eds Anne M. Cohler, Basia Carolyn Miller, and Harold Samuel Stone. Cambridge: Cambridge University Press.
- Narayan, Uma. 1997. *Dislocating Cultures: Identities, Traditions, and Third-World Feminism*. New York: Routledge.
- Navarro, Mireya. 2004. "The Most Private of Makeovers." *The New York Times*, (November 28), Section 9, 1–2.
- Nourse, Victoria. 1997. "Passion's Progress: Modern Law Reform and the Provocation Defense." *Yale Law Journal* 106 (March): 1331–448.
- Okin, Susan Moller. 1989. *Justice, Gender, and the Family*. New York: Basic Books.
- Okin, Susan Moller. 1998. "Feminism and Multiculturalism: Some Tensions." *Ethics* 108 (July): 661–84.
- Okin, Susan Moller. 1999. *Is Multiculturalism Bad for Women?* eds Joshua Cohen, Matthew Howard, and Martha C. Nussbaum. Princeton, NJ: Princeton University Press, 9–24.
- Oliver, Myrna. 1988. "Immigrant Crimes: Cultural Defense—Legal Tactic." *L.A. Times* (July 15), 29.
- Ortner, Sherry B. 1984. "Theory in Anthropology Since the Sixties." *Comparative Studies in Society and History* 26: 126–66.
- Ortner, Sherry B. 1996. *Making Gender: The Politics and Erotics of Culture*. Boston: Beacon Press.
- Parekh, Bhikhu. 2000. *Rethinking Multiculturalism. Cultural Diversity and Political Theory*. Cambridge, MA: Harvard University Press.
- Phillips, Anne. 2003. "When Culture Means Gender: Issues of Cultural Defence in the English Courts." *Modern Law Review* 66 (July): 510–31.
- People v. Chen*, No. 87-7774, Supreme Court, NY County [Dec. 2, 1988].
- People v. Hundal*, No. F037541 [Cal. App. 5 Dist. 2002].
- People v. Kobayashi*, No. B157685 [Cal. App. 2 Dist. 2003].
- People v. Mayberry*, 542 P.2d 1337 [Cal. 1975].
- People v. Moua*, No. 315972-0, Fresno County Superior Court [Feb. 7, 1985].
- Pequignot v. Detroit*, 16 F. 211 [1883].
- Perry v. Commonwealth*, 839 S.W.2d 268 [KY 1992].
- Rennison, Callie Marie. 2003. *Intimate Partner Violence, 1993–2001*. Washington D.C.: U.S. Department of Justice, Bureau of Justice Statistics, Crime Data Brief No. NCJ 197838.
- Renteln, Alison. 2004. *The Cultural Defense*. Oxford: Oxford University Press.
- Resnik, Judith. 1989. "Dependent Sovereigns: Indian Tribes, States, and the Federal Courts." *University of Chicago Law Review* 56: 671–759.
- Reynolds v. United States*, 98 U.S. 145 [1879].
- Rodebaugh v. State*, No. 436, 1990 WL 254365 [Del. 1990].
- Rosenblum, Nancy. 1997. "Democratic Sex: Reynolds v. U.S., Sexual Relations, and Community." In *Sex, Preference, and Family: Essays on Law and Nature*. Oxford: Oxford University Press, 63–85.
- Santa Clara Pueblo v. Martinez*, 436 U.S. 49 [1978].
- Schulhofer, Stephen J. 1998. *Unwanted Sex: The Culture of Intimidation and the Failure of Law*. Cambridge, MA: Harvard University Press.
- Shachar, Ayelet. 2001. *Multicultural Jurisdictions. Cultural Differences and Women's Rights*. Cambridge: Cambridge University Press.
- Sherman, Rorie. 1989. "Cultural Defenses Draw Fire." *National Law Journal* (April 17), 28.
- Sherman, Spencer. 1986. "When Cultures Collide." *California Lawyer* 6 (January), 36.
- Simpson, Audra. 2000. "Paths Toward a Mohawk Nation: Narratives of Citizenship and Nationhood in Kahnawake." In *Political Theory and the Rights of Indigenous Peoples*, eds Duncan Ivison, Paul Patton, and Will Sanders. Cambridge: Cambridge University Press, 113–36.
- Smith, Rogers M. 1997. *Civic Ideals: Conflicting Visions of Citizenship in U.S. History*. New Haven, CT: Yale University Press.
- Spinner-Halev, Jeff. 2001. "Feminism, Multiculturalism, Oppression, and the State." *Ethics* 112 (October): 84–115.
- Stanton, Elizabeth Cady. [1871] 1970. *Address at The Decade Meeting on Marriage and Divorce. In History of the National Woman's Rights Movement*, ed Paulina Wright Davis. New York: Source Book Press.
- Stanton, Elizabeth Cady, Susan B. Anthony, and Matilda Joselyn Gage, eds. [1881–1922] 1970. Reprint. *History of Woman Suffrage*, v. 1, 1848–1861 and v. 3, 1876–1885. New York: Source Book Press.
- State v. Little*, 462 A.2d 117 [N.H. 1983].
- State v. Reams*, 616 P.2d 498 [Or. 1980].
- State v. Wood*, 545 A.2d 1026 [Conn. 1988].
- U.S. v. Contreras*, 180 F.3d 1204 [1999].

- U.S. v. Guzman*, 236 F. 3d 830 [2001].
- U.S. v. Natal-Rivera*, 879 F.2d 391 [1989]
- Volpp, Leti. 1994. "(Mis)identifying Culture Asian Women and the 'Cultural Defense.'" *Harvard Women's Law Journal* 17 (Spring): 57–97.
- Volpp, Leti. 1996. "Talking 'Culture': Gender, Race, Nation, and the Politics of Multiculturalism." *Columbia Law Review* 96 (October): 1573–617.
- Volpp, Leti. 2001. "Feminism versus Multiculturalism." *Columbia Law Review* 101 (June): 1181–218.
- Waldron, Jeremy. 1995. "Minority Cultures and the Cosmopolitan Alternative." In *The Rights of Minority Cultures*, ed. Will Kymlicka. Oxford: Oxford University Press, 93–119.
- Warner, Michael. 2002. "Beyond Gay Marriage." In *Left Legalism/Left Critique*, eds Wendy Brown and Janet Halley. Durham: Duke University Press, 259–89.
- Wedeen, Lisa. 2002. "Conceptualizing Culture: Possibilities for Political Science." *American Political Science Review* 96 (December): 713–28.

# To Trust an Adversary: Integrating Rational and Psychological Models of Collaborative Policymaking

WILLIAM D. LEACH *California State University, Sacramento*

PAUL A. SABATIER *University of California, Davis*

*This study explores how trust arises among policy elites engaged in prolonged face-to-face negotiations. Mirroring recent evidence that citizens' procedural preferences (as opposed to policy preferences) drive trust in government, we find that interpersonal trust among stakeholders in consensus-seeking partnerships is explained by the perceived legitimacy and fairness of the negotiation process more so than by the partnership's track record of producing mutually agreeable policies. Overall, hypotheses derived from social psychology do as well or better than those based on rational-choice assumptions. Important predictors of trust include small and stable groups, generalized social trust, clear decision rules, political stalemate, congruence on policy-related beliefs, and absence of devil-shift (the belief that one's opponents wield more power than one's allies). Surprisingly, null or negative correlations exist between trust and network density, measured by membership in voluntary associations. The study illustrates the value of behavioral models that integrate institutional, rational, and psychological explanations.*

How can interpersonal trust be cultivated among policy elites engaged in long-term face-to-face negotiations? This article seeks to identify the precursors of interpersonal trust in high-stakes disputes where opposing sides have long histories of animosity and differ on fundamental values and perceptions. What are the beliefs and personal circumstances that predispose one policy actor to trust another? What are the institutional arrangements that foster trust during protracted multiparty policy deliberations? How can policymakers break the vicious cycle of distrust and noncooperative behavior and initiate a "virtuous cycle" of trust and cooperation (Putnam 2000)?

Understanding how policy elites build mutually trusting relationships is crucial in several contexts, with a unifying theme being that all politics is personal. Resolving conflicts through political means usually boils down to cooperation among two or more individual persons who, though they are political adversaries, must eventually achieve a rough consensus on disputes over policy or procedure. These face-to-face negotiations can last many months or years, such that personal relationships are likely to evolve over the course of negotiations, and can have a major influence on whether

an agreement is achieved. In legislatures, for example, harmonious interpersonal relationships help members work across party lines to move legislation through multiple veto points (Price 2004, Reingold 1996). Distrust, especially if unfounded, can result in gridlock in which legislators fail to achieve agreements or compromises that could benefit the constituents of both political parties. Interpersonal relationships also come into play during regulatory negotiations (Weber 1998) in which agencies work with interested parties to avoid protracted litigation, and during litigation itself, when parties usually prefer to negotiate out-of-court settlements rather than risk the expense and uncertainty of trial. In the context of international peace negotiations, distrust between negotiators can have particularly stark consequences in terms of life and death (Tyler 2000).

Although interpersonal trust is not always essential for achieving cooperation or collective action, it is an important catalyst in a wide range of policy-making contexts (Cook, Hardin, and Levi 2005). In an age where Congress is more partisan than ever (Hetherington 2001, 622) and where views on moral issues are increasingly polarized among those citizens who self-identify as political partisans (Evans 2003), understanding how trust arises among political actors is as important as ever.

Previous attempts to explain interpersonal trust have generally been guided by one of two prominent traditions in political science: institutional rational choice and social psychology. Both traditions view trust as a precursor to consensus building and collective action, but they diverge on how trust arises. Simply put, institutional rational choice views trust as resulting from recent evidence of the trustworthiness of other parties and from institutional rules that encourage trustworthiness; social psychology views *distrust* as arising from belief conflict, cognitive limitations, and misgivings about the legitimacy of the policymaking process. The roots of each tradition's differing predictions lie in their distinct models of individual behavior, to be outlined in the following.

William D. Leach is Research Director, Center for Collaborative Policy, California State University, Sacramento, 1303 J Street, Suite 250, Sacramento, CA 95814 (bleach@ccp.csus.edu).

Paul A. Sabatier is Professor, Department of Environmental Science & Policy, University of California, One Shields Avenue, Davis, CA 95616-8576 (pasabatier@ucdavis.edu).

We thank Mark Lubell and Chris Weible for suggesting several improvements to the manuscript, and Neil Pelkey for alerting us to the literature on interest-based negotiation and its relevance to the Advocacy Coalition Framework. The research was carried out by the authors and other members of the Watershed Partnerships Project at UC Davis: Maryann Hulsman, Steve Kropp, Tamara LaFramboise, Christal Love, Neil Pelkey, Jim Quinn, Kate Reza, Martha Turner, and Chris Weible. Funding was provided by the National Science Foundation's "Decision Making and Valuation for Environmental Policy" program (Grant 9815471), the United States Environmental Protection Agency's "Science to Achieve Results" (STAR) program (Grant R82-7145), the David and Lucile Packard Foundation's "Conserving California Landscapes Initiative," and the William and Flora Hewlett Foundation's Conflict Resolution Program.

On the premise that science progresses best by comparing and integrating the explanatory power of multiple theories rather than testing hypotheses drawn from a single theory alone (Allison 1969; Platt 1964; Stinchcombe 1968), this article considers both traditions simultaneously. Our goal is not to discard one theory or the other, or to rehash the well-documented limitations of rational choice (Green and Shapiro 1994); rather, it is to integrate insights from both traditions and to work toward defining the range of political situations where rationality drives trust versus those where social psychology dominates.

We consistently employ an inclusive and multi-dimensional definition of trust, as is advocated in much of the literature (e.g., Braithwaite 1998, 51–52; McKnight and Chervany 1996; Tonkiss and Passey 1999). As elaborated in the following, trust involves faith or confidence in another's propensity to keep promises, to negotiate honestly, to show respect for other points of view, and to express some level of concern for the welfare of others.

## INSTITUTIONAL RATIONAL CHOICE

In its simplest form, the rational choice model of the individual assumes a self-interested welfare maximizer whose ability to make optimal choices is curtailed mainly by imperfect information. The choice in question is whether to trust the other parties to the dispute—for example, whether to accept at face value the declarations or proposals they offer during the course of the negotiation. The decision to trust is made largely on the basis of information about the parties' past behavior in similar circumstances, and information about their incentives going forward that might influence whether they continue to negotiate in good faith and keep their promises or ultimately defect. Naturally, trust is higher among parties that have a history of reaching agreements and implementing their provisions. Agreements among adversaries demonstrate willingness to negotiate in good faith and to accept reasonable compromises. Successful implementation demonstrates a propensity to honor commitments and an ability to work competently. Trust ought to be correlated with the length, depth, and recency of past collaboration.

The decision to trust a policy adversary also involves assessing their incentives to cooperate or defect in the future. Institutional rational choice scholars view these incentives as being shaped largely by the presence of rules governing the negotiation, monitoring, and enforcement of consensual agreements (Hardin 2002, 127; Rothstein 2000; Ruscio 1999; Williamson 1993). The presence of such institutions enhances each individual's ability to make a credible commitment (Ostrom 1992, 302). Formalized collective choice rules specify how deliberations are to be conducted and decisions made. Such rules reduce opportunities for misunderstandings regarding the terms of an agreement or whether an agreement was reached. Monitoring rules provide confidence that people who break agreements

will be detected. Enforcement rules increase the probability that detected defectors will be punished. Taken together, such institutional rules ought to promote trust by discouraging antisocial behavior.

Several rational choice scholars have attempted to soften the assumption of self-interest by allowing for the possibility that cultural norms predispose individuals to trust others and behave in a trustworthy manner. For example, Ostrom's (1998; 13; 1999) work on common pool resources subsumes many of the hypotheses from social capital theory (Putnam 1993) regarding the positive feedback among specific interpersonal trust, generalized trust, and horizontal social networks. Generalized trust refers to perceptions about people at large, whereas interpersonal (a.k.a. specific) trust describes one's perceptions about specific individuals. One locks the door when leaving home for lack of generalized trust in passers by; one leaves a spare key with a neighbor as an expression of specific trust. Individuals with strong norms of generalized trust are more likely to place confidence in specific individuals (Putnam 2000).

Social networks build trust by providing opportunities for successful collective action. The strength of each interpersonal relationship ought to increase with the frequency of contact and with the cumulative number of interactions over time. Experimental evidence shows that face-to-face communication can lead to social capital, including trust (Ostrom, Gardner, and Walker 1994). Correlational evidence suggests that trust learned in one social circle often spills over to one's relationships beyond that circle (Putnam 1993, 174; 2000). If so, then policy elites who participate in a soccer team, sorority, and service club might be expected to express higher levels of trust toward their policy adversaries. Individuals should express stronger trust in their policymaking opponents if they interact with them frequently, if they have done so for many months, and if they themselves participate in a large number of unrelated voluntary associations.

Assessing the trustworthiness of others is more feasible when the number of parties to a policy dispute is relatively small and stable (Olson 1965). Likewise, the higher levels of surveillance and lack of anonymity that characterize small and stable communities encourage trustworthy behavior, leading to a reputation for trustworthiness. A good reputation is important because it gives others the confidence to speak openly and to provide favors with the expectation that a comparable gesture will be returned at some point in the future. Community stability also reduces each individual's discount rate, increasing their willingness to incur immediate costs to achieve delayed or long-term benefits of collaboration (Ostrom 1990, 35). People who plan to exit a policy arena in the near future have less incentive to invest in building constructive working relationships.

Softening the rationality assumptions can confound or even reverse certain hypotheses regarding institutions and trust. When a society with strong norms of trust and cooperation is subjected to a strict enforcement regime, the norms can weaken, resulting



in diminished cooperation overall (Lubell and Scholz 2001). Similarly, in cooperative games, trust and trustworthiness can be highest when enforcement is weak (Bohnet, Frey, and Huck 2001). When enforcement is strong, players trust the legal system to deter breaches of contracts, but they don't necessarily trust each other. When enforcement is virtually absent, players alter their behavior and *become* more trustworthy to attract lucrative contract offers.

Another confounding factor is the reciprocal nature of cause and effect between institutions and trust. Institutional rational choice theory predicts that trust should follow the adoption of suitable institutions, but social capital theorists reason that societies adopt these institutions only after trust is found to be insufficient to spur cooperation.<sup>1</sup> If institutions can be viewed as both a precursor to trust and a societal response to distrust, then institutions and trust might correlate either positively or negatively in a cross-sectional study. A positive correlation would indicate that the rational-choice mechanism dominates. A negative correlation would indicate that the social capital mechanism is stronger. A null correlation would be ambiguous; either both effects are absent or both are equal in magnitude and cancel out.

## SOCIAL PSYCHOLOGY

Political scientists have long been interested in models of individual behavior that depart from the economist's assumptions of self-interested rationality. The fields of cognitive and social psychology, in particular, have generated several useful insights related to cognitive dissonance (Festinger 1957), biased assimilation (Lord, Ross, and Lepper 1979; Munro and Ditto 1997), computational constraints (Simon 1985), risk aversion (Quattrone and Tversky 1988; Tversky and Kahneman 1981), and belief system hierarchies (Converse 1964; Lakatose 1971). Rarely, however, have scholars attempted to spell out what these assumptions imply for the dynamics of trust among political elites engaged in policy negotiations. Two exceptions are the Advocacy Coalition Framework developed by political scientists Sabatier and Jenkins-Smith (1993, 1999) and the literature on interest-based negotiation, which grew out of scholarship in law and business administration (e.g., Fisher and Ury 1981).

Both the Advocacy Coalition Framework and the interest-based negotiation literature assume that an individual's policy-relevant beliefs are nested in a hierarchy. At the highest level are "core beliefs" or "underlying interests," which are relatively general in scope and difficult to change. At a lower level are "secondary beliefs" or "policy positions," which are relatively narrow in scope and malleable.

The fundamental insight from the interest-based negotiation literature is that the classic horse-trading

model of negotiation, in which each party concedes one or more policy positions to achieve a compromise agreement, frequently leads to socially suboptimal solutions that serve the fundamental interests of neither party. An interest-based approach, by contrast, employs a Tocquevillian notion of enlightened self-interest in which the parties agree to invest time and energy inventing novel policy proposals that address not only one's own underlying interests but also those of one's adversaries. Thus, interest-based negotiation requires the parties to publicly reveal sufficient information about their own core interests to allow the other parties to work toward policy proposals that might satisfy them. To pursue an interest-based negotiated settlement is an inherently trusting act. It requires the parties to recognize that, although they might not share the same underlying interests as their policy adversaries, their adversaries' interests are still legitimate and worthy of being satisfied. Interest-based negotiation requires faith in the willingness of others to negotiate honestly and without malice. It also requires faith in the basic fairness of the collaborative policymaking process. Each individual's level of interpersonal trust should therefore correlate with both their enthusiasm for consensus-based decision making and the perceived fairness of a given collaborative process.

The Advocacy Coalition Framework employs a belief hierarchy to help explain how individuals assess the trustworthiness of other parties. Specifically, the framework suggests that individuals assess trustworthiness by comparing their own core beliefs to those of other parties. Relative to secondary beliefs, core beliefs ought to be the most efficient guides to the trustworthiness of others because they are more general in scope than secondary beliefs. Within the core, the Framework discriminates between deep core<sup>2</sup> and policy core<sup>3</sup> beliefs. Because policy core beliefs are more directly salient to specific policy disputes than are deep core beliefs, the Framework hypothesizes that "the policy core provides the principle glue of coalitions" (Zafonte and Sabatier 1998), and the principle foundation of trust. Reliance on heuristic indicators of trustworthiness is necessary because each individual's ability to process and analyze information is assumed to be limited by time and computational constraints (Simon 1985), making it unfeasible to systematically evaluate the other parties' past behavior and institutional incentives (the focus of institutional rational choice theory).

<sup>2</sup> Deep core beliefs are fundamental, normative, and ontological axioms that operate across all policy sectors. Examples include the prioritization of competing values (e.g., freedom vs. security), the proper scope of government authority, and the relationship between people and nature.

<sup>3</sup> Policy core beliefs identify the welfare tradeoffs deemed appropriate for achieving deep core values within a policy sector. Examples include judgments about the relative importance of competing social groups (e.g., tribal vs. commercial fishermen), and the relative importance of competing problems (e.g., scarcity of water for environmental vs. agricultural uses).

<sup>1</sup> Putnam (2000, 145) charts the explosive growth in the per capita number of police, security guards, lawyers, and judges over the last 40 years, and concludes that Americans have increasingly invested in "the rule of law" to compensate for declining social capital.

Consistent with the literature on cognitive dissonance and biased assimilation, the Advocacy Coalition Framework assumes that preexisting beliefs strongly influence the filtering of new information, especially at the policy core level (Lord, Ross, and Lepper 1979; Munro and Ditto 1997). Individuals who differ on core beliefs see the world through different lenses and often interpret a given piece of evidence in different ways. This produces distrust because people who reach opposite conclusions on factual issues tend to question each other's motives or reasonableness. Even on policy topics where the scientific evidence is relatively clear, policy elites who lack a common set of perceptual filters will tend to view their adversaries as backward, ignorant, or malevolent. The problem is exacerbated when the relevant data are ambiguous.

If belief conflict is a major source of distrust, then building trust requires a convergence of beliefs over time. Sabatier and Zafonte (2001) suggest that such a convergence (termed "policy-oriented learning across coalitions") is only feasible for disputes that are analytically tractable (meaning accepted quantitative data and theory exist) and when the conflict is mediated through a "professional forum" in which a neutral facilitator forces scientific experts from competing coalitions to justify their claims before their peers using accepted standards of data quality and inference. Achieving agreement on empirical issues should enhance trust by demonstrating that opponents are in fact reasonable people who can be convinced by sound evidence.

The Advocacy Coalition Framework borrows prospect theory's assumption of risk aversion, meaning people weigh losses more heavily than gains (Quattrone and Tversky 1988; Tversky and Kahneman 1981). Risk aversion in combination with the filtering of new information implies that policy elites will remember their political defeats more vividly than their victories. As a result, policy actors tend to view their opponents as being more powerful than they actually are, a phenomenon termed "devilshift" (Sabatier, Hunter, and McLaughlin 1987). Individuals experiencing devil shift distrust their opponents even more because they perceive their opponents as having the means to cause harm—not just the will. Moreover, people who are risk averse probably approach new interpersonal relationships with great caution and skepticism, not unconditional trust. Trust should be inversely related to devil shift.

Finally, the interest-based negotiation literature predicts that a political stalemate is a necessary condition for a policy dispute to be ripe for negotiation (Fisher 1997; Kriesberg 1998). In other words, the parties must mutually perceive that their best alternative to negotiation (usually litigation, lobbying, or the status quo) is unlikely to produce satisfactory results. If none of the parties can successfully pursue their agenda in an alternate venue such as the courts or legislature, then they need not worry about other parties defecting from the negotiations by appealing to external authorities. A mutual stalemate would bolster each party's confidence that the other parties will respect the consensus-based process rather than switch venues.

## STAKEHOLDER PARTNERSHIPS AS POLICYMAKING VENUES

Stakeholder partnerships are one type of collaborative policymaking forum in which trust is thought to be critical for success, and where building trust is often an explicit, instrumental goal (Leach and Pelkey 2001). Stakeholder partnerships consist of policy elites who convene about once a month to discuss or negotiate public policy within a broadly defined issue area. Most partnerships include representatives from private advocacy groups, local governments, and state and federal agencies. The primary goal is to achieve consensus regarding formal agency rulemaking, discretionary agency actions, or voluntary commitments from the private sector. Unlike other forms of consensus-oriented policymaking, partnerships are intended to last several years as they address multiple interrelated issues of concern to various stakeholders (Leach, Pelkey, and Sabatier 2002).

Stakeholder partnerships are increasingly common, particularly in the field of natural resource and environmental policymaking. Since the late 1980s, partnerships have formed throughout the United States to address disputes over water quality and related land use issues (Kenney 1999). The U.S. Environmental Protection Agency (EPA) has catalogued 3,500 partnerships and other watershed groups nationwide (EPA 2002).

Given their popularity and potential influence on environmental outcomes, stakeholder partnerships are an excellent setting for studying the evolution of interpersonal trust among policy elites engaged in long-term, face-to-face negotiations. Understanding how they function is a worthy task for political scientists interested in institutional design, regulation, devolution, interest groups, or environmental policy.

## METHODS

The study examines 76 stakeholder partnerships dealing with local watershed policy and implementation, randomly sampled from the states of California and Washington. Quantitative case studies of the 76 partnerships were compiled between 1999 and 2003. The field research began by identifying all partnerships in California that were active at any point between 1995 and 2000, including partnerships defunct at the time of the study. To be included in the sampling frame, a partnership needed to meet at least four times per year and focus on managing one or more streams, rivers, or watersheds. To ensure adequate diversity of stakeholders, each partnership needed to include (1) at least one state or federal official; (2) at least one representative of local government—either a general-purpose city or county, or a special district (such as water or school district); and (3) at least two opposing interests, such as a resource user and either a regulatory agency or an environmental group.

The search revealed a population of 150 partnerships in California, from which 47 were randomly sampled with geographic stratification, such that no more than two partnerships were selected from a single

watershed.<sup>4</sup> In Washington we randomly selected 20 watersheds and sampled 1 to 2 partnerships from each.<sup>5</sup> Because the selection process was random and the sample size is relatively large, the overall results should be representative of watershed partnerships in the two states.

The sample includes 12 partnerships that had disbanded by the time of the study. In three cases, they had achieved their main objectives. The other nine disbanded after their negotiations ended in stalemate.

For each selected partnership, we first interviewed three to six key participants including the partnership's coordinator/facilitator plus at least one key participant from a pro-environment perspective and at least one participant from a pro-development perspective. Second, we analyzed relevant documents such as watershed plans and meeting minutes. Finally, we mailed a survey to all participants in the partnership plus several stakeholders who were knowledgeable about the partnership but were not members.

For the survey, names of participants and nonparticipant stakeholders were obtained during the interviews. The smallest partnership had six survey recipients, and the largest had 76. Of 2,498 surveys, 1,625 were returned at least partially completed, yielding a response rate of 65%.<sup>6</sup> Response rates for individual partnerships ranged from 45% to 88%.

Throughout the paper, the unit of analysis is the individual survey respondent. The models of interpersonal trust presented in the following section include individual-level variables from the survey (e.g., respondent's social network density) and partnership-level variables gleaned from the interviews and documents (e.g., the number of participants). The construction of each variable is described briefly in the following section. Details are given in Appendix B. Appendix A (Table 2) presents descriptive statistics for each variable.

### Measuring the Dependent Variable: Interpersonal Trust

The survey measures interpersonal trust by asking each respondent, with respect to their own partnership, "How many of the participants (a) are honest, forthright, and true to their word? (b) have reasonable motives and concerns? (c) are willing to listen, and sincerely try to understand other points of view? (d) reciprocate acts of good will or generosity? (e) propose solutions that are compatible with the needs of other members of the partnership?" Respondents answer each question by indicating: 1 = none, 2 = few, 3 = half, 4 = most, 5 = all. The five questions are averaged to generate a composite measure of trust. As such, the

scale measures the *breadth*, rather than the *intensity*, of each respondent's trust in other members of the partnership. Focusing on breadth of trust is appropriate in the context of consensus-based policymaking, where distrust of any single participant can derail agreements or other forms of collective action by the partnership as a whole.

The composite scale is reliable and consistent as indicated by a high Chronbach's alpha (0.87) and the high Pearson's correlations between each question and the scale (respectively, 0.78, 0.77, 0.84, 0.84, and 0.80). The high interitem correlations lend empirical support to our decision to employ a composite, multidimensional measure of trust rather than define trust more narrowly.<sup>7</sup>

The data on interpersonal trust are normally distributed with a mean of 3.6 and a standard deviation of 0.65. In other words, two thirds of respondents believe that at least half—but less than all—of the participants are trustworthy.

### Measuring Rational Choice Explanatory Variables

Measures for each variable are detailed under Appendix B. In the following, is an overview of how each variable is operationalized.

Three variables capture aspects of the stakeholders' breadth and success of past collaboration. First, each stakeholder is characterized as pertaining to a *defunct partnership* or an active one, as of the time of the study. Second, the *partnership age* is measured in months from inception to the time of the study, or to disbandment in the case of defunct partnerships. Third, to quantify *partnership agreements*, we use interview data and documents to devise an ordinal 5-point index where the top score indicates agreement on a comprehensive watershed management plan.

To characterize each partnership's institutional rules, interview data are coded to generate qualitative variables indicating the presence or absence of *deliberation groundrules*, *decision-making rules*, *compliance monitoring rules*, and *enforcement rules* (see Appendix B).

Three variables related to the likelihood of reputation for trustworthiness include (1) *partnership size*, the number of participants in the respondent's partnership, gleaned from interview data and documents; (2) *stable relationships*, the proportion of the other stakeholders with whom the respondent expects to continue interacting over the next five years; and (3) *nonparticipant observers*, who are distinguished from active partnership participants by asking survey respondents to self-identify into either category.

<sup>4</sup> We partitioned California using Hydrologic Unit Code watersheds defined by the U.S. Geological Survey. There are 160 watersheds in the state, ranging from 35 to 9,000 square miles.

<sup>5</sup> We partitioned Washington using the 62 Water Resource Inventory Areas, which range from 140 to 3,000 square miles.

<sup>6</sup> Assuming response rate definition, RR2 (The American Association for Public Opinion Research 2000, 36).

<sup>7</sup> In the regression model presented below, substituting any one of the five variables for the full trust scale yields results similar to those for the full scale, although the model fit is not as strong, as would be expected considering that scales generally reduce measurement error, and considering that the scale in question has high internal reliability (Chronbach's alpha 0.87).

*Generalized trust in people* and *generalized trust in public officials* are measured using questions replicated from the General Social Survey.

The *social network density* of each respondent is measured as a count of the number of voluntary associations the respondent affiliates with, patterned after a question from the General Social Survey.

### Measuring Social Psychology Explanatory Variables

Each respondent's enthusiasm for the consensus-building process is measured with two variables. First, the respondent's *consensus norm* is assessed by asking whether "consensus-based negotiation among stakeholders, including agencies," is a valuable approach for resolving disputes, when juxtaposed with three alternatives: agency purview, landowner purview, or tradable permits. Second, respondents are asked to assess the *procedural fairness* of the process in terms of whether it "treats all parties fairly and consistently."

*Deep core belief conflict* and *policy core belief conflict* compare the respondent's beliefs to the average beliefs within the partnership. The greater the difference, the higher the respondent will score on each belief conflict variable. Deep core beliefs are measured using a five-question scale regarding the respondent's *laissez-faire* conservatism. Policy core beliefs are measured by asking respondents to evaluate the relative seriousness of 13 problems in the watershed, ranging from degraded water quality to weakened property rights.

*Net devil shift* (devil shift minus angel shift) is measured by asking respondents to name the three most powerful or influential members of the partnership, the three members they consider their closest allies, and their three main opponents. Devil shift is the proportion of powerful members who are opponents. Angel shift is the proportion who are allies.

*Mutual stalemate* is proxied by asking each respondent whether they lack alternate venues outside the partnership, suitable for pursuing their interests. Responses from within each partnership are averaged to create a partnership-level variable measuring the extent to which parties mutually perceive that their alternatives to negotiation are limited.

## RESULTS AND DISCUSSION

What are the institutional and psychological factors that bolster trust among policy adversaries, and what is the relative importance of each? Table 1 presents three linear regression models of interpersonal trust; one using rational choice variables, one using social psychology variables, and one combined model. Each model employs ordinary least-squares regression<sup>8</sup> with

<sup>8</sup> Multicollinearity was checked three ways. First, variance inflation factors in all models are low, never exceeding 2.1. Second, no model exhibits both a high condition index (i.e., >15) plus two or more variables with high variance proportions (i.e., >0.8). Finally, randomly sampling 50% of the observations and rerunning the analysis yields similar results

robust Huber-White standard errors.<sup>9</sup> All three models fit the data reasonably well, with adjusted  $R^2$  statistics of 0.21 for the rational choice model, 0.25 for the social psychology model, and 0.34 for the combined model. Of the 1,625 total observations, the proportion dropped due to listwise deletion for missing data in the three models is 16%, 13%, and 20%, respectively.

### The Rational Choice Model of Trust

In the institutional rational choice model, seven of the 13 coefficients are both statistically significant and correlated with trust in the predicted direction.<sup>10</sup> One coefficient (that for social network density) is statistically significant ( $p = 0.03$ ), but *inversely* related to trust, contrary to the prediction. When the social psychology variables are added to the model, the coefficient becomes insignificant ( $p = 0.43$ ). A related finding is that social network density, which was measured as a count of the voluntary associations in which the respondent participates, is unrelated to *generalized* trust in people ( $r = 0.02$ ) or public officials ( $r = 0.01$ ) among the sampled policy elites. These results are puzzling in light of several studies that found strong correlations between membership in voluntary associations and trust among the general public (e.g., Brehm and Rahn 1997; Putnam 2000; Putnam, Leonardi, and Nanetti, 1993).

Other studies, however, failed to find any relationship, or found nonlinear relationships between trust and voluntary associations (National Commission on Civic Renewal 1998; Sullivan and Transue 1999). In their review of this literature, Hibbing and Theiss-Morse (2002, 186) conclude that voluntary groups "all too often do nothing to help people learn to come to a democratic solution on divisive issues" because most groups are too homogenous (Peel 1998) or shun controversial issues (Eliasoph 1998) or promote "trust of those you know or distrust of those you do not" (Levi 1996, 49).

We speculate that another mechanism is operating in our own study, which looks exclusively at policy elites, whose social networks are quite dense relative to the general public.<sup>11</sup> If most stakeholders are extremely well networked, then joining a few more voluntary associations would presumably add little additional capacity for trust. That is, the positive effect of networks on generalized trust may eventually plateau. To explain the inverse relationship, we speculate that at very high levels, network density reflects social capital that

<sup>9</sup> In the event that the errors are correlated for respondents within a partnership, ordinary least-squares regression would underestimate the error variance. The Huber-White standard errors, when clustered on a variable identifying the partnership each respondent belongs to, are consistent even if the errors are not identically distributed (Moulton, 1986)

<sup>10</sup> Dropping nonsignificant variables from the model generates little impact on the remaining coefficients or the  $R^2$

<sup>11</sup> Respondents in the present study reported membership in an average of 6.2 voluntary associations (median = 5), whereas the general public reports an average of 1.8 associations (median = 1), as reflected by the General Social Survey, Question 328, survey years 1974-1994

**TABLE 1. Models of Interpersonal Trust**

Independent Variables	Institutional Rational Choice Model		Social Psychology Model		Combined Model	
	B	$\beta$	B	$\beta$	B	$\beta$
Partnership age <sup>a</sup>	0.000 (0.000)	0.00			-0.000 (0.000)	-0.01
Partnership defunct <sup>a</sup>	-0.237** (0.085)	-0.12			-0.165 (0.089)	-0.08
Policy agreements reached <sup>a</sup>	0.057 (0.042)	0.07			0.017 (0.030)	0.02
Deliberation groundrules <sup>a</sup>	-0.027 (0.070)	-0.02			0.011 (0.039)	0.01
Decisionmaking rules <sup>a</sup>	0.218** (0.075)	0.10			0.125* (0.056)	0.06
Compliance monitoring rules <sup>a</sup>	0.024 (0.079)	0.02			-0.011 (0.057)	-0.01
Enforcement rules <sup>a</sup>	0.062 (0.061)	0.04			0.055 (0.063)	0.03
Partnership size <sup>a</sup>	-0.006*** (0.001)	-0.17			-0.004** (0.001)	-0.11
Stable relationships	0.115*** (0.018)	0.20			0.092*** (0.015)	0.16
Knowledgeable outside observer	-0.155** (0.044)	-0.09			-0.123** (0.041)	-0.07
Generalized trust in people	0.252*** (0.049)	0.16			0.239*** (0.046)	0.15
Generalized trust in officials	0.195*** (0.039)	0.14			0.102** (0.035)	0.07
Social network density	-0.009* (0.004)	-0.06			-0.003 (0.004)	-0.02
Deep core belief conflict			-0.078** (0.028)	-0.09	-0.058* (0.028)	-0.06
Policy core belief conflict			-0.012*** (0.003)	-0.13	-0.011*** (0.003)	-0.12
Net devil shift			-0.187*** (0.043)	-0.13	-0.162*** (0.039)	-0.11
Consensus norm			0.088*** (0.013)	0.22	0.081*** (0.013)	0.20
Procedural fairness <sup>a</sup>			0.239*** (0.032)	0.25	0.147*** (0.033)	0.15
Mutual stalemate <sup>a</sup>			0.116** (0.040)	0.09	0.086* (0.040)	0.07
Constant	2.811*** (0.112)		1.745*** (0.240)		1.778*** (0.247)	
Adjusted $R^2$	0.21		0.25		0.34	
Number of cases	1373		1415		1293	

Note: OLS regression with unstandardized (B) and standardized ( $\beta$ ) coefficients, and in parentheses, Huber White robust standard errors with clustering on a variable identifying each respondent's partnership. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

<sup>a</sup> Variables that are constant for all respondents within a partnership, but vary across partnerships

is spread too thin, leaving individuals with insufficient time or energy to develop strong trusting relationships with other members of the partnership, *on average*. In other words, once generalized trust plateaus, adding additional nodes to one's social network simply reduces the proportion of individuals whom one trusts highly.

Judging by the relative size of the standardized coefficients, the most important correlates of trust in the rational choice model are the number of participants in the partnership (inversely related to trust), stable interpersonal relationships, and generalized trust in people and government. The importance of generalized trust suggests that some stakeholders were already

predisposed to be trusting when they first joined the partnership. This baseline generalized trust seems to influence the eventual formation of specific trust among partnership members. Causation in the opposite direction is less likely because one's specific experiences within the partnership should have only a marginal effect on general attitudes, which are shaped by the sum of all experience. For example, Hetherington's (1998) reciprocal model of trust in elected officials suggests that general attitudes affect specific ones "much more powerfully than the reverse."

Of the four categories of institutional rules examined (deliberation groundrules, decision-making rules,

compliance-monitoring rules, and enforcement rules), only the presence (vs. absence) of decision-making rules is statistically significant. Specifically, interpersonal trust is lower in partnerships that have not yet determined the rules for making collective choices (e.g., simple majority vote, unanimity, and informed consent) or where the participants expressed confusion about whether such rules were in place. This result lends support to the rational choice hypothesis that clear decision rules bolster confidence in the ability of the other participants to make credible commitments.

The insignificance of monitoring and enforcement rules is difficult to interpret because rational choice theorists view such rules as precursors to trust, whereas social capital theorists view the adoption of such rules as a societal response to low trust. The null correlations might indicate that both mechanisms are equal in magnitude and cancel out in a cross-sectional analysis. Alternatively, both mechanisms might actually be absent, especially considering that consensus-based policy negotiations, like international courts, involve parties that participate on their own volition. In such arrangements where participants voluntarily submit to the authority of the collective body, moral suasion typically plays as large a role in enforcing agreements as does the threat of formal sanctions (Rieke and Kenney 1997, 51). Moreover, legal impediments prevent many of the parties in watershed partnerships, such as public agencies and sovereign tribes, from formally subordinating themselves to the partnership. Accordingly, monitoring and enforcement rules occurred with relatively low frequency among the partnerships sampled (32% and 18%, respectively).

The model includes three variables that capture aspects of the partnerships' breadth and success of past collaboration: the partnership's age, its status as active or defunct, and the level of agreement achieved to date. Only active status correlates with trust in the rational-choice model, but it is insignificant in the combined model. The age of the partnership and the level of agreement achieved are significant in neither model. These results imply that incremental gradations of success are not predictive of trust, but the abject failures and severed relationships represented by defunct partnerships do predict distrust, at least prior to controlling for the social psychology variables.

The model also includes a binary variable distinguishing partnership participants from other local stakeholders who did not self-identify as participants. The variable is statistically significant, indicating that participants in partnerships display more trust than do knowledgeable outside observers.<sup>12</sup>

<sup>12</sup> We designed the study to include relevant non-participant stakeholders partly to guard against selection bias. For example, the absence of effective institutional rules conceivably could have caused such a loss of trust that these potential participants selected out of the partnership (Tyler and McGraw 1986). To test whether the institutional factors affected nonparticipants differently than participants, we added variables for the interaction effect between participant status and each of the partnership-level variables (those denoted by superscript "a" in Table 1). In the resulting model (not shown) none of the 10 interaction terms are statistically significant, and the direct

## The Social Psychology Model of Trust

Each of the six social psychology variables correlates significantly with trust in the predicted manner in both the reduced and the combined models. The two strongest correlates of trust in both the reduced and the combined models are the stakeholders' general confidence in the legitimacy of consensus-based decision making, and their confidence in the fairness of their specific collaborative process.

The analysis shows that trust is inversely related to belief conflict on representative measures of both deep core ideology (*laissez-faire* conservatism) and policy core beliefs (judgments about the relative seriousness of various problems within a policy sector). The Advocacy Coalition Framework had predicted that policy core conflict should impair trust more so than deep core conflict. The coefficients for each variable are consistent with this prediction in both the reduced and the combined models, but the difference between coefficients is not statistically significant in the reduced model ( $p = 0.12$ ). In summary, the models suggest that disagreement over the relative importance of various problems is at least as important as disagreement over whether government should play a liberal or conservative role in regulating economic activity.

Both the reduced and the combined models support the importance of two of the more innovative variables from the social psychology literature. First, net devil shift correlates negatively with the degree of trust. The more one views opponents as being more powerful than allies, the lower the degree of trust in partnership participants. Second, both the reduced and combined models indicate that stakeholders trust one another more when they collectively lack opportunities to undercut the partnership by shopping for alternate venues.

## PRACTICAL IMPLICATIONS FOR COLLABORATIVE POLICYMAKING

What are the implications for policy elites engaged in collaborative policymaking? Some of the variables identified previously are more easily manipulated than others. For example, core beliefs and generalized trust are probably very difficult to change over the short term. The best that a mediator or facilitator can do is call attention to their importance and hope that participants will muster their most productive attitudes. The alternative—limiting participation to those who hold moderate policy views and endorse consensus decision

effects of each variable are essentially unchanged from those in the simple models of Table 1. In conclusion, we find little evidence that the institutional variables affect participants and nonparticipants differently. However, a remaining question is whether the suppressed levels of trust among nonparticipants caused them to select out, or whether selecting out buffered them from the trust-building effects of the collaborative process, or both. Nonparticipants tend to have slightly weaker consensus norms (5.6 vs. 5.8 for participants, on a 7-point scale,  $p < 0.01$ ) and slightly less positive perceptions of the fairness of particular partnerships (4.5 vs. 5.0,  $p < 0.01$ ), but again it is unclear whether these views caused or resulted from non-participation.



making—would have the added benefit of reducing group size, but could undermine trust by damaging the perceived legitimacy of the process (Leach and Pelkey 2001). The importance of procedural fairness suggests that facilitators should periodically assess whether participants feel respected and perceive themselves as having an appropriate degree of control over the outcomes of the negotiations (Tyler and Blader 2000). The mere presence of procedural groundrules—a statistically insignificant variable from the rational choice model—is apparently not sufficient to bolster trust.

One of the clearest findings is that trust is higher among participants who plan to interact with the other members of the partnership over the next 5 years. Partnership conveners could bolster trust by making it clear at the outset that the process is designed to last several years into the future. Watershed partnerships typically take about 48 months to reach formal agreements and to implement restoration, education, or monitoring projects (Leach, Pelkey, and Sabatier 2002).

This analysis may be the first to empirically link trust to devil shift—the perception that one's opponents are more powerful than one's allies. Devil shift appears to be an important facet of the syndrome of distrust among policy elites. The implication of this finding depends on whether the perceived power imbalance is accurate or exaggerated. If the perception is accurate, facilitators should pay particular attention to bolstering trust among politically weaker parties. If the power imbalance is imaginary, facilitators could design exercises to lead stakeholders toward accurate assessments of their relative power within the partnership by exploring each party's best available alternative to negotiation. One of the purposes of adopting a consensus-based decision rule, as most partnerships do, is to help level the balance of power within the partnership itself.

The importance of policy core belief conflict also has implications for how negotiations are structured. If distrust stems from disagreement over which problems are most serious, then deliberations should begin with a period of "joint fact finding" and consensus building on the basic dimensions of the various problems. Partnerships can also pursue empathy-building exercises, such as field trips to the local businesses or environmental sites affected by the partnership's actions. Over the long term, partnerships can commission scientific investigations to settle disputed facts.

Another finding with potential implications for institutional design is the association between trust and mutual political stalemate, measured as a collective lack of alternatives to negotiation. In a political system like that of the United States where legal authority is diffuse, alternative venues will always be present. What counts, however, is the extent to which participants in the partnership view them as viable (Baumgartner and Jones 1993). Although partnerships have little ability to curtail their members' rights to appeal decisions through the legal system, legislators and agency administrators can help limit routes of appeal by signaling their commitment to respect decisions reached through consensus within a partnership.

## METHODOLOGICAL AND THEORETICAL IMPLICATIONS: INTEGRATING SOCIAL PSYCHOLOGY AND RATIONAL CHOICE

This study illustrates the utility of testing two or more models at a time, and where appropriate, seeking to integrate models under one unified framework. Because data are often consistent with multiple theories, testing hypotheses from a single theory often leads to overconfidence (Stinchcombe 1968). By testing multiple models, one can ascertain which fits the data best, or whether two or more models are complementary, each contributing predictive power. Testing multiple models also helps researchers avoid two common biases: confirmation bias (a tendency to seek confirming evidence) and theory tenacity (persistent belief in a theory despite contrary evidence; Loehle 1987). By investing professional and emotional energy in at least two theories, researchers buffer themselves against the inevitable psychological stress that occurs when empirical results contradict the predictions of any one theory.

Another argument for multiple models is their necessity for strong inference, the term Platt (1964) coined for conclusions drawn from experimental studies that conclusively discriminate between two or more competing hypotheses. Platt argued persuasively that strong inference is the most efficient path to progress in science. However, the comparative strategy cannot produce strong inference until theory is relatively mature, with clear and precise predictions (Loehle 1987, 399). In fields where theory is relatively imprecise or phenomena are especially complex or studies are difficult to devise, strong inference remains a laudable yet unattainable ideal. Even so, comparative studies are valuable when they demonstrate that emerging models perform comparably to more mainstream models and deserve to be cultivated further.

Over the last 20 years, political scientists have predominantly pursued single-theory studies and have often grounded their models of the individual in microeconomics and its rationality assumptions. The results presented earlier suggest the discipline should renew its ties to other traditions such as social and cognitive psychology. The Institutional Analysis and Development framework (Ostrom 1999) represents one such attempt to relax the assumption of rationality by allowing for cultural norms and other community characteristics that predispose individuals to behave in socially desirable ways. In fact, our findings suggest that the classic rational choice variables (i.e., institutional incentives and evidence of past trustworthiness) have relatively little influence on interpersonal trust in the context of long-term policy negotiations. Only one of the four institutional rules variables in the model is significantly correlated with trust.

In this data set, the bulk of explanatory power from the rational choice variables comes from those variables borrowed from social capital theory. Specifically, interpersonal trust is highest among individuals in small groups with stable relationships and strong norms of generalized trust in people and government. Similarly, the significance of stakeholders' perceptions of

procedural fairness and consensus norms, when juxtaposed with the statistical insignificance of institutional groundrules, lends support to Ostrom's (1999) increasing emphasis on norms over rules. We hasten to add that institutional rules might contribute to perceptions of fairness, but evidence of such a mechanism is mixed at best in this dataset.<sup>13</sup> By taking rationality assumptions as the starting point for modeling individual behavior, political scientists may overemphasize the importance of institutional rules.

Social capital theory makes important contributions to the rational choice model, but the results also firmly contradict the social capital hypothesis that trust results from network density. At least within the realm of policy elites in watershed partnerships, the number of voluntary associations participants belong to correlates with trust *inversely* or not at all, depending on how the model is specified.

By contrast, all six hypotheses derived from social psychology are supported. The paramount importance of stakeholders' perceptions of the fairness and legitimacy of the process (the variables *consensus norms* and *procedural fairness*) parallels the finding of Hibbing and Theiss-Morse (2001, 2002) that people's attitudes toward government are driven primarily by their satisfaction with *how* government operates, not by its track record of producing agreeable policies. Just as policy space (the distance between citizen's preferred policies and those the government actually produces) explains trust in government to a lesser extent than process space, so too are stakeholders' perceptions of procedural fairness and legitimacy better predictors of *interpersonal* trust than is the partnership's track record of producing policy agreements. Policy space does come into play in our model of interpersonal trust, but only at the level of policy-related *values* (the variable *policy core beliefs*), not the partnership's success in actually forging policy agreements. Specifically, trust declines with the distance between each individual's policy-related values and the average values of other members of the partnership. In sum, the social psychological emphasis on process norms and core values explains trust better than the rational emphasis on hard evidence of trustworthiness, as revealed by the parties' ability to compromise and reach formal agreements on policy.

With the social psychology model performing comparably or better than the rational choice model in terms of individual hypotheses and overall fit, the social psychology model clearly merits attention in future research. Nonetheless, the goal is not necessarily

to discard one framework or the other. After many years of trying, the authors have never identified any truly opposing hypotheses between the two frameworks, where one predicts a positive correlation between two variables, and the other predicts a negative correlation. Instead, when explaining the development of trust or collective action, each framework emphasizes a distinct and sometimes overlapping set of explanatory variables. Although the two frameworks are rooted in separate branches of the social sciences and are built around strikingly different assumptions, they are not incompatible. Political actors are at once rational and psychological creatures. Ideally, models of political behavior would integrate these two human faces.

One strategy for pursuing such an integrated framework is to further define the scope of the two underlying models (Loehle 1987, 401). For example, scholars could seek to identify the range of political situations where rationality dominates human behavior, and those that call forth the psyche. The findings of this study suggest that a more complex model incorporating insights from social psychology is particularly useful when:

- the policymaking process is being conducted through prolonged face-to-face deliberations
- the format of the process is relatively novel (such as collaborative, consensus-based processes)
- the relative influence of various actors is ambiguous (thus feeding the devil shift)
- stakeholders disagree over fundamental values and norms
- the issues are scientifically complex, such that policy actors also disagree on the relative severity and causes of different problems
- monitoring and enforcement mechanisms are difficult or impossible to establish (such as negotiations among autonomous and highly heterogeneous stakeholders, as in the present study)

We speculate that rationality may dominate in situations where it is easier to calculate the probability of defection, or where the stakeholders have more direct financial stakes in the outcome of negotiations. If true, this would parallel the well-documented observation that self-interest drives citizens' policy preferences only when the personal costs and benefits of a policy are highlighted (Young et al. 1991) or are especially clear (Chong, Citrin, and Conley 2001) or substantial (Green and Gerken 1989). Further research would be required to test these propositions about rational and psychological roots of trust.

Given the prospect of profitable collaboration on the one hand, and costly betrayal on the other, to trust a political adversary is a weighty and complex decision. Trust is an elusive phenomenon—part emotion and part rational calculation—part reflex and part deliberate act. Social scientists seeking to identify the determinants of trust in political contexts will meet with greater success if they are willing to consider multiple theoretical frameworks and employ models of human behavior that are suitably elaborate.

<sup>13</sup> Individual-level variation in perceived procedural fairness correlates moderately well with several institutional variables: partnership age ( $r = .22$ ), partnership defunct ( $r = -.28$ ), policy agreements reached ( $r = .28$ ), decision-making rules ( $r = .18$ ), and compliance monitoring rules ( $r = .24$ ). On the other hand, none of these institutional variables is statistically significant if procedural fairness is substituted for trust as the dependent variable in the rational choice model. Instead, perceived fairness is correlated with stable relationships and generalized trust, and inversely related to non-participant observer status and social network density ( $N = 1310$ , adjusted  $R^2 = 0.14$ ).

## APPENDIX A. DESCRIPTIVE STATISTICS

TABLE 2. Descriptive Statistics and Correlations with Interpersonal Trust

Variables	<i>r</i>	<i>N</i>	Min.	Max.	Mean	Std. Dev.
Interpersonal trust	1.00	1531	1	5	3.64	0.647
Partnership age <sup>a</sup>	0.10***	1625	4	332	66.9	50.33
Partnership defunct <sup>a</sup>	-0.20***	1625	0	1	0.13	0.33
Policy agreements reached <sup>a</sup>	0.13***	1625	2	5	3.64	0.851
Deliberation groundrules <sup>a</sup>	-0.02	1625	0	1	0.745	0.436
Decision-making rules <sup>a</sup>	0.12***	1625	0	1	0.895	0.306
Compliance-monitoring rules <sup>a</sup>	0.18***	1625	0	1	0.346	0.476
Enforcement rules <sup>a</sup>	0.05*	1625	0	1	0.184	0.388
Partnership size <sup>a</sup>	-0.14***	1625	6	100	25.2	16.88
Stable relationships	0.29***	1444	1	5	3.29	1.103
Knowledgeable outside observer	-0.10***	1625	0	1	0.220	0.414
Generalized trust in people	0.24***	1562	0	1	0.78	0.42
Generalized trust in officials	0.21***	1563	0	1	0.66	0.47
Social network density	-0.04	1582	0	61	6.15	4.35
Deep core belief conflict	-0.20***	1586	0	4	1.01	0.725
Policy core belief conflict	-0.22***	1568	5.28	51	22.9	7.105
Net devil shift	-0.22***	1542	-1	1	-0.060	0.448
Consensus norm	0.31***	1585	1	7	5.74	1.61
Procedural fairness <sup>a</sup>	0.35***	1625	3.13	7	4.870	0.649
Mutual stalemate <sup>a</sup>	0.21***	1625	3.71	7	4.818	0.512

Note \* $p < .05$ , \*\*\* $p < .001$  (two-tailed Pearson's correlation).

<sup>a</sup> Variables that are constant for all respondents within a partnership, but vary across partnerships.

## APPENDIX B: VARIABLE DEFINITION AND SURVEY QUESTIONS

Data are drawn from the survey unless noted otherwise. Wording of survey questions is indicated by quotation marks. Unless noted, questions are close-ended 7-point scales where 1 = strongly disagree, 7 = strongly agree; and 9 = no opinion.

**Partnership age.** Months from inception to the time the interviews were conducted, or until disbandment in the case of defunct partnerships.

**Policy agreements reached.** Coded by the research team from interviews and partnership documents. (1 = no agreement, 2 = agree on which issues to discuss, 3 = agree on general goals or principles, 4 = agree on one or more restoration projects, 5 = agree on a comprehensive watershed management plan.)

**Deliberation groundrules.** Dummy variable coded by the research team from interviews and partnership documents. Codeform wording: "Were formal process or groundrules established?"

**Decision-making rules.** Dummy variable coded by the research team from interviews and partnership documents. Codeform wording: "Decision rules: None or decision rules haven't been settled yet."

**Compliance monitoring rules.** Dummy variable coded by the research team from interviews and partnership documents. Codeform wording: "Did the partnership monitor compliance with agreements?"

**Enforcement rules.** Dummy variable coded by the research team from interviews and partnership documents. Codeform wording: "Were sanctions used in cases of noncompliance?"

**Partnership size.** The average number of people in attendance at partnership meetings. Coded by the research team from interviews and partnership documents.

**Stable relationships.** "Please indicate whether the following statements apply to none, few, half, most, or all

of the participants in the partnership. How many of the participants... do you expect to keep interacting with over the next five years?" (5-point scale: 1 = none, 2 = few, 3 = half, 4 = most, 5 = all).

**Knowledgeable outside observer.** "Do you consider yourself a participant in the Partnership?" (1 = Yes, 0 = No. Recoded to 1 = outsider, 0 = participant.)

**Generalized trust in people.** "Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?" (1 = Would take advantage, 2 = Would try to be fair, 3 = Other. Recoded to: 1 = Would try to be fair, 0 = Would take advantage or Other.)

**Generalized trust in public officials.** "Most public officials (people in public office) are not really interested in the problems of the average person." (1 = Agree, 2 = Disagree, 3 = Other. Recoded to 1 = Agree, 0 = Disagree or Other.)

**Social network density.** Modeled after the 1987 General Social Survey (questions 328–356), this variable tallies the responses to the following 11-part question. "For each category below, please tell us how many different groups you participate in. For example, if you are a member of a softball team and a chess club, write '2' on the first line (a) Recreational clubs: sports teams, hobby clubs, birding groups, etc. (b) Religion-affiliated groups or congregations. (c) Youth groups. (d) Culture or ethnicity groups. (e) Service organizations. (f) Fraternity or sorority. (g) Veterans groups. (h) Business or professional associations, or labor unions. (i) Property rights groups. (j) Environmental advocacy groups. (k) Other organizations."

**Deep core belief conflict.** The absolute value of the difference between the respondent's *laissez-faire* conservatism and the mean level of *laissez-faire* conservatism within the respondent's partnership. *Laissez-faire* conservatism is a scale calculated as the mean of the following five questions (a) The best government is the one that governs the least. (b) A first consideration of any good political system is the

protection of private property rights. (c) Government laws and regulations should primarily ensure the prosperity of business because the health of the nation is dependent upon the well-being of business. (d) Government planning almost inevitably results in the loss of essential liberties and freedoms. (e) Decisions about development are best left to the economic market." (Chronbach's alpha 0.82; Spearman's correlations between each question and the scale: 0.78, 0.81, 0.76, 0.79, 0.65, respectively).

**Policy core belief conflict.** Using respondents' perceptions of the relative seriousness of 13 problems in the watershed, this variable is an index of the (absolute value) difference between the respondent's perceptions and the average perception of all members of the partnership. Formally:

$$\frac{1}{13} \sum_{i=1}^{13} ABS(X_i - Y_i),$$

where  $X_i$  = Respondent's perceived seriousness of Problem  $i$ .  $Y_i$  = Mean perceived seriousness of Problem  $i$  within the respondent's partnership.

"Please indicate the current seriousness of the following problems for your watershed. Using the thermometer scale below, a score of 100 indicates an extremely serious problem, while a score of 0 indicates the issue is not a problem at all. Impaired water quality, Inadequate water supply, Lack of open space, Threats to species or habitat, Risk of catastrophic fire, Risk of damaging floods, Excessive gov't regulation or taxes, Threats to private property or water rights, Threats to tribal or treaty rights, Excessive population growth or urban development, Lack of economic prosperity, Conflict among stakeholders, Other key issue" (optional write-in).

**Allies.** "Please identify up to three organizations/interests that you regard as allies on issues important to the partnership."

**Opponents.** "Please indicate up to three organizations/interests that you disagree with most frequently on issues important to the partnership."

**Powerful.** "Please indicate the three organizations/interests that are most important or influential regarding partnership issues."

**Devil shift.** The proportion of powerful stakeholders who are opponents.

**Angel shift.** The proportion of powerful stakeholders who are allies.

**Net devil shift.** Devil shift minus angel shift.

**Consensus norm.** Subquestion A of the following: "Listed below are four alternative approaches for managing Watersheds. For each alternative, please circle the response that best reflects your opinion. 'The best strategies for resolving watershed issues include . . . (a) consensus-based negotiations among stakeholders, including agencies. (b) reliance on each agency's legal mandate and court review. (c) reliance on tradable permits for water, fish catch, development, etc. (d) allowing private property owners to manage their land as they see fit.'"

**Procedural fairness.** "The partnership process treats all parties fairly and consistently." Responses from each partnership are averaged to create a partnership-level indicator.

**Mutual stalemate.** "If the partnership fails to adopt workable solutions, my concerns could probably be satisfied by appealing directly to the legislature, courts, or individual agencies." The scale is reversed, and then responses from each partnership are averaged to create a partnership-level variable.

## REFERENCES

- Allison, Graham T. 1969. "Conceptual Models and the Cuban Missile Crisis." *American Political Science Review* 63 (September): 689–718.
- American Association for Public Opinion Research. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lenexa, Kansas: AAPOR.
- Baumgartner, Frank R., and Bryan D. Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.
- Bohnet, Iris, Bruno S. Frey, and Steffen Huck. 2001. "More Order with Less Law: On Contract Enforcement, Trust and Crowding." *American Political Science Review* 95 (March): 131–44.
- Braithwaite, Valerie. 1998. "Communal and Exchange Trust Norms: Their Value Base and Relevance to Institutional Trust." In *Trust and Governance*, ed. Valerie Braithwaite and Margaret Levi. New York: Russell Sage, 46–74.
- Brehm, John, and Wendy Rahn. 1997. "Individual-Level Evidence for the Causes and Consequences of Social Capital." *American Journal of Political Science* 41 (July): 999–1023.
- Chong, Dennis, Jack Citrin, and Patricia Conley. 2001. "When Self-Interest Matters." *Political Psychology* 22 (September): 541–70.
- Converse, Phillip. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David Apter. New York: The Free Press, 206–61.
- Cook, Karen S., Russell Hardin, and Margaret Levi. 2005. *Cooperation Without Trust?* New York: Russell Sage.
- Elasoph, Nina. 1998. *Avoiding Politics*. Cambridge, MA: Cambridge University Press.
- EPA (U.S. Environmental Protection Agency). 2002. *Adopt Your Watershed*. Accessed 29 August 2002. <http://www.epa.gov/adopt>
- Evans, John H. 2003. "Have Americans' Attitudes Become More Polarized?—An Update." *Social Science Quarterly* 84 (March): 71–90.
- Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row Peterson.
- Fisher, Ronald J. 1997. *Interactive Conflict Resolution*. Syracuse, NY: Syracuse University Press.
- Fisher, Roger, and William Ury. 1981. *Getting to Yes: Negotiating Agreement without Giving In*. Markham, Ontario: Penguin Books.
- Green, Donald Philip, and Ann Elizabeth Gerken. 1989. "Self-Interest and Public Opinion Toward Smoking Restrictions and Cigarette Taxes." *The Public Opinion Quarterly* 53 (Spring): 1–16.
- Green, Donald P., and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale University Press.
- Hardin, Russell. 2002. *Trust and Trustworthiness*. New York: Russell Sage.
- Hetherington, Marc J. 1998. "The Political Relevance of Political Trust." *The American Political Science Review* 92 (December): 791–808.
- Hetherington, Marc J. 2001. "Resurgent Mass Partisanship: The Role of Elite Polarization." *American Political Science Review* 95 (September): 619–31.
- Hubbing, John R., and Elizabeth Theiss-Morse. 2001. "Process Preferences and American Politics: What the People Want Government to Be." *American Political Science Review* 95 (March): 145–53.
- Hubbing, John R., and Elizabeth Theiss-Morse. 2002. *Stealth Democracy. Americans' Beliefs about How Government Should Work*. Cambridge, MA: Cambridge University Press.
- Kenney, Doug S. 1999. "Historical and Sociopolitical Context of the Western Watersheds Movement." *Journal of the American Water Resources Association* 35 (June): 493–503.
- Kriesberg, Louis. 1998. *Constructive Conflicts. From Escalation to Resolution*. Lanham, MD: Rowman and Littlefield.
- Lakatos, Imre. 1971. "History of Science and Its Rational Reconstruction." In *In Memory of Rudolf Carnap*. Vol. 8 of *Boston Studies in the Philosophy of Science*, ed. R. C. Buck and Robert S. Cohen. Dordrecht, Holland: D. Reidel, 42–134.
- Leach, William D., and Neil W. Pelkey. 2001. "Making Watershed Partnerships Work: A Review of The Empirical Literature." *Journal of Water Resources Planning and Management* 127 (November/December): 378–85.

- Leach, William D., Neil W. Pelkey, and Paul A. Sabatier. 2002. "Stakeholder Partnerships as Collaborative Policymaking. Evaluation Criteria Applied to Watershed Management in California and Washington." *Journal of Policy Analysis and Management* 21 (Fall): 645–70.
- Levi, Margaret. 1996. "Social and Unsocial Capital: A Review Essay of Robert Putnam's 'Making Democracy Work'." *Politics and Society* 24 (March): 44–55.
- Loehle, Craig. 1987. "Hypothesis Testing in Ecology: Psychological Aspects and the Importance of Theory Maturation." *Quarterly Review of Biology* 62 (December): 397–409.
- Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37: 2098–2109.
- Lubell, Mark, and John T. Scholz. 2001. "Cooperation, Reciprocity, and the Collective-Action Heuristic." *American Journal of Political Science* 45 (January): 160–78.
- McKnight, D. Harrison, and Norman L. Chervany. 1996. "The Meaning of Trust." Working Paper. Carlson School of Management, University of Minnesota.
- Moulton, Brent R. 1986. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*, 32 (August): 385–97.
- Munro, Geoffrey, and Peter Ditto. 1997. "Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information." *Personality and Social Psychology Bulletin* 23 (June): 636–53.
- National Commission on Civic Renewal. 1998. "Associations and Democracy." In *A Nation of Spectators. How Civic Disengagement Weakens America and What We Can Do About It*. [http://www.puaf.umd.edu/Affiliates/CivicRenewal/finalreport/associations\\_and\\_democracy.htm](http://www.puaf.umd.edu/Affiliates/CivicRenewal/finalreport/associations_and_democracy.htm) (June 22, 2004).
- Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.
- Orbell, John, and Robyn M. Dawes. 1991. "A 'Cognitive Miser' Theory of Cooperators' Advantage." *American Political Science Review* 85 (June): 515–28.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, MA: Cambridge University Press.
- Ostrom, Elinor. 1992. "The Rudiments of a Theory of the Origins, Survival, and Performance of Common-Property Institutions." In *Making the Commons Work: Theory, Practice, and Policy*, ed. Daniel W. Bromley. San Francisco: Institute for Contemporary Studies Press.
- Ostrom, Elinor A. 1998. "Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997." *American Political Science Review* 92 (March): 1–22.
- Ostrom, Elinor. 1999. "Institutional Rational Choice: An Assessment of the Institutional Analysis and Development Framework." In *Theories of the Policy Process*, ed. Paul Sabatier. Boulder, CO: Westview Press.
- Ostrom, Elinor, Roy Gardner, and James Walker. 1994. *Rules, Games, and Common-Pool Resources*. Ann Arbor, MI: The University of Michigan Press.
- Peel, Mark. 1998. "Trusting Disadvantaged Citizens." In *Trust in Governance*, ed. Valerie Braithwaite and Margaret Levi. New York: Russell Sage.
- Platt, John R. 1964. "Strong Inference." *Science* 146: 347–53.
- Popper, Karl. 1959. *The Logic of Scientific Discovery*. New York: Harper and Row.
- Price, David E. 2004. *The Congressional Experience*. Third Edition. Boulder, CO: Westview Press.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
- Putnam, Robert D., with Robert Leonardi and Raffaella Y. Nanetti. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Quattrone, George, and Amos Tversky. 1988. "Contrasting Rational and Psychological Analyses of Political Choice." *American Political Science Review* 82 (September): 719–36.
- Reingold, Beth. 1996. Conflict and Cooperation: Legislative Strategies and Concepts of Power among Female and Male State Legislators. *The Journal of Politics* 58 (2, May): 464–85.
- Rieke, Betsy, and Douglas Kenney. 1997. *Resource Management at the Watershed Level: An Assessment of the Changing Federal Role in the Emerging Era of Community-Based Watershed Management*. Boulder, CO: Natural Resources Law Center, University of Colorado School of Law.
- Rothstein, Bo. 2000. "Trust, Social Dilemmas and Collective Memories." *Journal of Theoretical Politics* 12 (October): 477–501.
- Ruscio, Kenneth P. 1999. "Jay's Prouette, or Why Political Trust Is Not the Same as Personal Trust." *Administration & Society* 31 (November): 639–57.
- Sabatier, Paul, Susan Hunter, and Susan McLaughlin. 1987. "The Devil Shift: Perceptions and Misperceptions of Opponents." *Western Political Quarterly* 40 (September): 51–73.
- Sabatier, Paul, and Hank Jenkins-Smith. 1993. *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder. Westview Press.
- Sabatier, Paul, and Hank Jenkins-Smith. 1999. "The Advocacy Coalition Framework: An Assessment." In *Theories of the Policy Process*, ed. Paul Sabatier. Boulder, CO: Westview Press.
- Sabatier, Paul A., and Matthew A. Zafonte. 2001. "Policy Knowledge: Advocacy Organizations." In *International Encyclopedia of the Social & Behavioral Sciences*, ed. Paul B. Baltes and Neil J. Smelser. New York: Elsevier Science Ltd, 11,563–11,568.
- Simon, Herbert. 1985. "Human Nature in Politics: The Dialogue of Psychology with Political Science." *American Political Science Review* 79 (June): 293–304.
- Stinchcombe, Arthur L. 1968. *Constructing Social Theories*. University of Chicago Press.
- Sullivan John L., and John E. Transue. 1999. "The Psychological Underpinnings of Democracy: A Selective Review of Research on Political Tolerance, Interpersonal Trust, and Social Capital." *Annual Review of Psychology* 50 (February): 625–50.
- Tonkiss, Fran, and Andrew Passey. 1999. "Trust, Confidence and Voluntary Organizations: Between Values and Institutions." *Sociology* 33 (May): 257–74.
- Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211: 453–58.
- Tyler, Tom R. 2000. "Social Justice. Outcome and Procedure." *International Journal of Psychology* 35 (April): 117–25.
- Tyler, Tom R., and Steven Blader. 2000. *Cooperation in Groups: Procedural Justice, Social Identity, and Behavioral Engagement*. Philadelphia: Psychology Press.
- Tyler, Tom R., and Kathleen M. McGraw. 1986. "Ideology and the Interpretation of Personal Experience. Procedural Justice and Political Quiescence." *Journal of Social Issues* 42 (Summer): 115–28.
- Weber, Edward P. 1998. "Successful Collaboration. Negotiating Effective Regulations." *Environment* 40 (November): 10–21.
- Williamson, Oliver E. 1993. "Calculativeness, Trust, and Economic Organization." *Journal of Law and Economics* 36 (April): 453–86.
- Young, Jason, Cynthia J. Thomsen, Eugene Borgida, John L. Sullivan, and John H. Aldrich. 1991. "When Self-Interest Makes a Difference: The Role of Construct Accessibility in Political Reasoning." *Journal of Experimental Social Psychology* 27 (May): 271–96.
- Zafonte, Matthew, and Paul Sabatier. 1998. "Shared Beliefs and Imposed Interdependencies as Determinants of Ally Networks in Overlapping Subsystems." *Journal of Theoretical Politics* 10 (October): 473–505.

# Using Geographic Information Systems to Study Interstate Competition

WILLIAM D. BERRY *Florida State University*

BRADY BAYBECK *University of Missouri—St. Louis*

**S**cholars have proposed two distinct explanations for why policies diffuse across American states: (1) policymakers learn by observing the experiences of nearby states, and (2) states seek a competitive economic advantage over other states. The most common empirical approach for studying interstate influence is modeling an indicator of a state's policy choice as a function of its neighbors' policies, with each neighbor weighted equally. This can appropriately specify one form of learning model, but it does not adequately test for interstate competition: when a policy diffuses due to competition, states' responses to other states vary depending on the size and location of specific populations. We illustrate with two substantive applications how geographic information systems (GIS) can be used to test for interstate competition. We find that lottery adoptions diffuse due to competition—rather than to learning—but find no evidence of competition in state choices about welfare benefits. Our empirical approach can also be applied to competition among nations and local jurisdictions.

**T**here is abundant evidence that public policies diffuse across the American states (e.g., Berry and Berry 1990, Mooney and Lee 1995). But why are one state's policymakers influenced by the policy choices of other states? Several explanations have been proposed, and two of the most common—policy learning and economic competition—reflect fundamentally different policymaking processes (Boehmke and Witmer 2004).

Some scholars maintain that states are influenced by the policy choices of other states because policymakers learn from the experiences of other states (e.g., Glick and Hays 1991, Mooney and Lee 1995). When confronted with a problem, decision makers simplify the task of finding a solution by choosing an alternative that has proven successful elsewhere (e.g., Simon 1997, Walker 1969). Most scholars who identify learning as the cause of interstate influence argue that diffusion of policy tends to be regional, with states looking primarily toward their neighbors or other nearby states for policies to emulate. Proximate states tend to share cultural, socioeconomic, and political characteristics that make them excellent "laboratories" for observing the likely effect of a policy choice (e.g., Walker).<sup>1</sup>

Other researchers attribute the diffusion of policy to competition: state officials make policy choices to gain an economic advantage over proximate states. They compete to attract perceived "goods" (e.g., businesses, affluent taxpayers) and to deter perceived "bads" (e.g., loss of tax revenue, immigration of poor persons) (e.g., Bailey and Rom 2004, Ka and Teske 2002).

Regardless of whether scholars point to learning or competition (or both) to justify their models of interstate influence, the vast majority of empirical tests of such models have relied on a similar specification of influence—one that assumes that states are affected equally by all their neighbors, and unaffected by more distant states. For example, most recent tests of models of state diffusion have relied on event history analysis in which the dependent variable is the probability that a state not yet having a policy will adopt it, and one of the independent variables is the number (or proportion) of neighboring states that have previously adopted (e.g., Berry and Berry 1990, Ka and Teske 2002). It is predicted that a rise in the number of neighbors that have adopted a policy results in an increase in the probability of adoption.

When testing a learning model that emphasizes influence by proximate states, this specification of interstate influence—although simple—is quite reasonable, as states looking to nearby states for policy cues should be more likely to emulate a policy adopted by many neighbors than a policy adopted by few. However, the simple "number of neighbors" variable does not suffice for testing for the presence of economic competition. When interstate influence is due to competition, states' influences on each other should vary depending on the size and location of specific populations of individuals and firms within the states. For example, assume that the diffusion of lottery adoptions across states (observed by Berry and Berry [1990]) is due to competition, with states adopting lotteries for fear of losing revenues when residents travel to another state to play. If this were the case, a state that is surrounded by states with lotteries, but in which most people live over a hundred miles from a state border would be less

William D. Berry is Marian D. Irish Professor and Syde P. Deeb Eminent Scholar in Political Science, Department of Political Science, Florida State University, Tallahassee, FL 32306-2230 (wberry@fsu.edu).

Brady Baybeck is Assistant Professor, Department of Political Science and Program in Public Policy Administration, University of Missouri—St. Louis, 348A Social Sciences Building, 8001 Natural Bridge Road, Saint Louis, MO 63121 (baybeck@umsl.edu).

An earlier version of this paper was presented at the 2004 Meeting of the Midwest Political Science Association, Chicago, IL. We thank Charles Barrilleaux, Frances Berry, Justin Esarey, Richard Fording, David Kimball, David Lowery, Chris Mooney, Chris Reenock, Evan Ringquist, John Scholz, and Craig Nolden for their comments on earlier drafts. To save space, some tables of statistical results are excluded; all such tables are in an unpublished supplement available at the ICPSR Publication-Related Archive along with a replication data set.

<sup>1</sup> Not all learning models, however, are regional, as illustrated by Grossback, Nicholson-Crotty, and Peterson's (2004) model of diffusion among ideologically similar states.



likely to adopt a lottery than a state that has a single neighbor with a lottery, but most of its population in a city right on the border of that one state. Until recently, a model predicting this form of competition could not be tested because it was infeasible to measure the size and location of populations in all fifty states, especially over a multiyear period. But advances in geographic information systems (GIS) have made such measurements possible.

GIS have not seen wide use by political scientists (for exceptions, see Cho 2003 and Gimpel and Schuknecht 2003). However, we illustrate the value of GIS by applying them to test competition models of two state policy choices—the level of welfare benefits and the adoption of a lottery—for which scholars have posited that both learning and competition are responsible for policy diffusion. To allow a presentation of the applications in a relatively small amount of space, we modify models developed in earlier studies: Berry and Berry's (1990) model of state lottery adoptions, and Berry, Fording, and Hanson's (2003) welfare benefit model. Taken along with analyses of the original studies, which we contend are good tests for the presence of policy learning, our empirical analysis using GIS constitutes a critical test for whether the interstate influence detected by the original authors is due to learning or competition.

## MEASURING THE CONCERN BY STATE OFFICIALS THAT MOTIVATES INTERSTATE COMPETITION

Berry and Berry's [hereafter B&B] (1990) model of the diffusion of the lottery proposes that the probability that a state without a lottery will adopt one is positively related to the number of its neighbors that have one, as states without lotteries adopt one to prevent the loss of revenues that occurs when residents cross the border to play other states' lotteries. Berry, Fording, and Hanson's [hereafter, BFH] (2003) welfare benefit model predicts that states try to set their benefits below those in surrounding states to discourage poor residents of other states from moving for more generous assistance. These models, although characterizing different policies, share an implicit presumption about the motivation of state officials to compete with nearby states. Both assume that a policy choice of a state—whether to adopt a lottery, or the size of its welfare benefit—is driven by the degree of concern of state officials about some form of behavior undertaken by citizens: a state resident buying a lottery ticket in another state or a poor person moving from another state for more generous public assistance. If we could observe these levels of concern, we could test these assumptions empirically.

Both models assume that state legislators and the governor are the principal officials involved in setting policy (B&B 1990, BFH 2003). One way to measure the degree of concern of these officials would be to survey them. However, obtaining responses from current office holders in every state would be difficult, and

getting accurate assessments of the degree of concern of officials during previous legislative sessions would be virtually impossible. Lacking the ability to observe directly the level of concern of state officials, we use GIS to estimate *Degree of concern*. Although measurement details are different for the welfare and lottery applications, our basic strategy is the same:

- First, we identify the individuals at risk of engaging in the behavior of concern (buying a lottery ticket in another state, or moving to the state for better welfare benefits). We call these persons the *population of concern*. [For the lottery application, this consists of adults in the state living near a state with a lottery. For the welfare application, the population of concern of state A is all poor people who live (1) in other states with welfare benefits appreciably lower than state A's and (2) close to an appealing location in state A.]
- We then estimate, for each person in the population of concern, the individual's *propensity to engage in the behavior of concern*—a value assumed to be a function of geographic location. [For our lottery analysis, someone's propensity to buy a lottery ticket in another state is assumed to be inversely related to the distance of the person's residence from the nearest state with a lottery. For the welfare application, the propensity of a poor person from another state to move to state A for better welfare benefits is presumed to be a function of both the distance of the person's residence from an appealing location in state A and the difference between the benefit level in his or her current state and that available in state A.]
- Finally, given the estimated propensity to engage in the behavior of concern by individuals in the state's population of concern, the degree of concern of state officials is estimated by summing propensity values over all persons in the population of concern, and norming the sum by dividing by a measure of state population (for the lottery application, the state's adult population; for our welfare analysis, the state's poor population). This norming by state population is performed because the degree of concern reflected by any sum of individual propensity values should vary depending on the size of a state.

## DISTINCT LEARNING AND COMPETITION HYPOTHESES

B&B (1990, 403) defend their prediction that the probability that a state will adopt a lottery is positively related to the number of previously adopting neighboring states by pointing to both policy learning ("previous adoptions by nearby states... yield important information about a [lottery's] effects") and economic competition ("people living near the border... can cross [state lines] to purchase tickets"). We recognize that both learning and competition are plausible explanations for the diffusion of lottery adoptions, but offer distinct propositions consistent with the two explanations:

**Lottery Learning Hypothesis:** The probability that a state will adopt a lottery is positively related to the number of states bordering it that have previously adopted.

**Lottery Competition Hypothesis:** The probability that a state will adopt a lottery is positively related to the degree of concern of its officials about residents going to other states to play the lottery.

B&B's (1990) event history model of state lottery adoptions has as dependent variable whether a state without a lottery adopts one in a year and includes as independent variables a set of "internal determinants" (e.g., state fiscal health, proximity to a gubernatorial election) and the number of previously adopting neighboring states. B&B interpret the coefficient estimate for number of previously adopting states as a general test for the presence of regional diffusion (which they attribute to both learning and competition). In contrast, we view their empirical analysis as a test for the occurrence of policy learning, and fashion a distinct test for the presence of interstate competition by substituting measures of state officials' degree of concern about residents going to other states to play the lottery for the neighbors variable in B&B's model.

There have been many studies of interstate influence over welfare benefit levels (Bailey and Rom 2004; BFH 2003; Rom, Peterson, and Scheve 1998; Volden 2002). The vast majority of these have framed their empirical analysis as a test of the "race to the bottom" thesis: the supposition that state officials compete to keep their benefit levels below those in nearby states to discourage immigration by the poor. If the thesis is correct, one determinant of a state's benefit level should be the degree of concern of its officials about poor people moving to the state for better welfare benefits, a variable influenced not only by benefit levels in nearby states but also by the size and location of the poor population in these states. A few studies, however, have raised policy learning as an alternative explanation for states adjusting their benefits in response to their neighbors' changes in benefits (Allard 1998, Tweedie 1994). Setting a welfare benefit level is a difficult and controversial choice. Thus, policymakers may seek "benchmarks" for comparison, and benefit levels in neighboring states are an obvious and reasonable frame of reference. Consequently, when the benefit level in a state increases relative to the average benefit level in neighboring states, the state should decrease its benefit level in the following year (to bring it closer to that available in benchmark states). We believe that both learning and competition are plausible explanations for states' benefit levels being influenced by their neighbors', and we introduce hypotheses consistent with both explanations:

**Welfare Learning Hypothesis:** An increase in a state's welfare benefit relative to the average benefit in neighboring states prompts a decrease in the state's benefit in the following year.

**Welfare Competition Hypothesis:** An increase in the degree of concern by state officials about

poor people in other states moving to the state for better welfare benefits prompts a decrease in the state's benefit in the following year.

The dependent variable in BFH's (2003) model of welfare benefits is a state's real (i.e., inflation-adjusted) Aid to Families with Dependent Children (AFDC) benefit level. One of its independent variables is the state's AFDC benefit level relative to the average benefit level in neighboring states in the previous year. The authors maintain that the coefficient for "benefit relative to neighbors" indicates the strength of interstate benefit competition.<sup>2</sup> We view BFH's coefficient estimate as testing the welfare learning hypothesis. To test the welfare competition hypothesis, we substitute measures of state officials' concern about welfare migration for "benefit relative to neighbors" in BFH's model.

Thus, our strategy for testing the lottery and welfare competition hypotheses requires us to construct measures of the degree of concern by state officials (about residents going to other states to play the lottery, and about poor people in other states moving to the state for better welfare benefits). We now describe our use of GIS to accomplish this task.

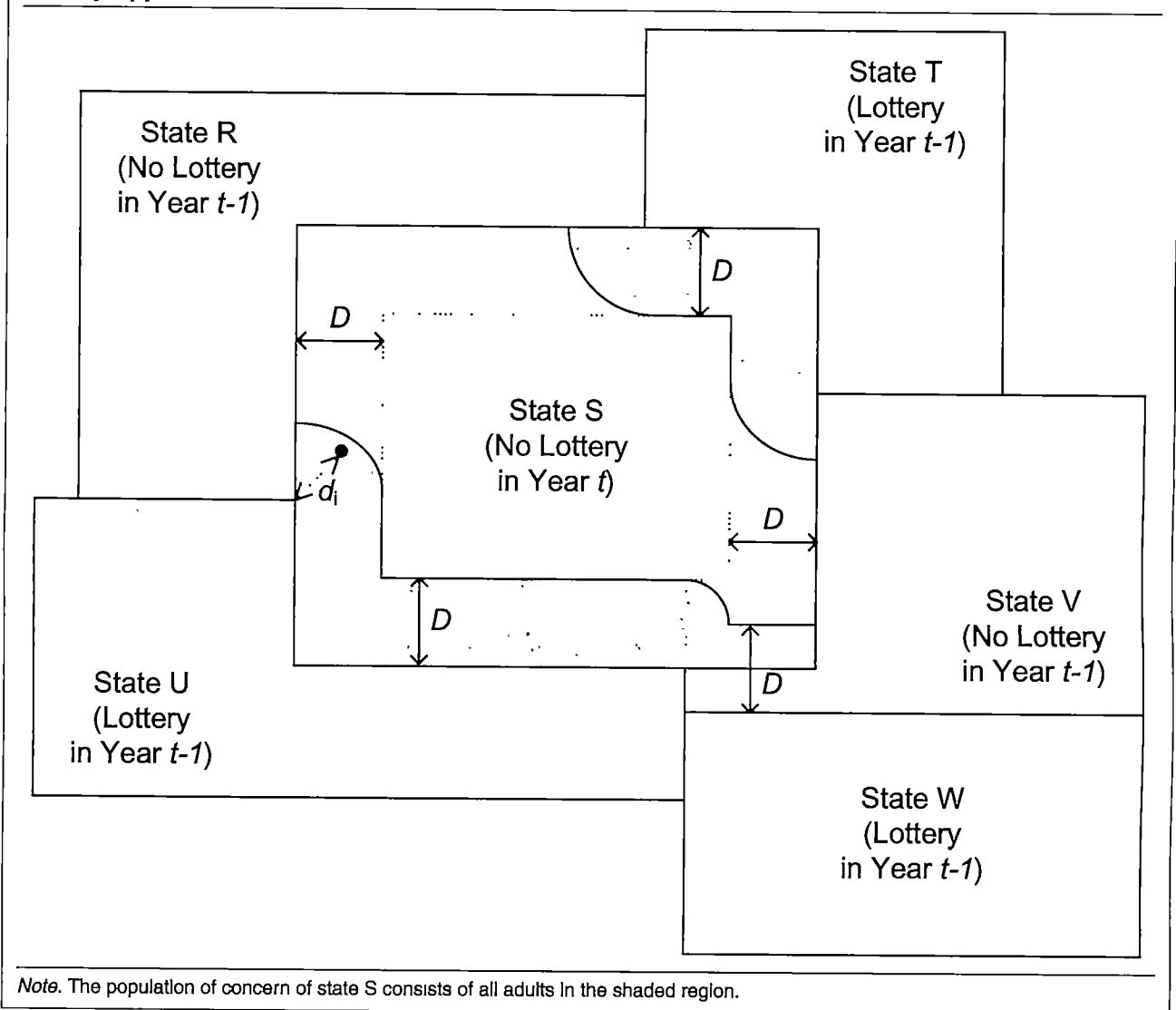
## USING GIS TO ESTIMATE DEGREE OF CONCERN ABOUT THE LOSS OF LOTTERY REVENUE

### Identifying the Population of Concern

Let  $D$  be the maximum distance (in miles) a person would be willing to travel to purchase a lottery ticket. Although  $D$  undoubtedly varies across individuals, to make our model tractable, we assume this value is constant.<sup>3</sup> Thus, for any state,  $s$ , without a lottery, we define the population of concern (those individuals at risk of going to other states to play the lottery) in a year as all adults in the state who live less than  $D$  miles from a state that had a lottery prior to the beginning of the year. The shaded region in Figure 1 depicts the population of concern of a hypothetical state  $[S]$  within  $D$  miles of five states, three of which have a lottery ( $T$ ,  $U$ , and  $W$ ) and two of which do not ( $R$  and  $V$ ). Note that to obtain the shaded region, we form a band of width  $D$  internal to the border of state  $S$ , but we exclude those sections of the band that are not within  $D$  miles of a state with a lottery. The most extreme southeast corner of state  $S$  is in the shaded region by virtue of being within  $D$  miles of nearby (but non-neighboring) state  $W$ .

<sup>2</sup> Most other studies of benefit competition also measure welfare benefits in real dollars (e.g., Bailey and Rom 2004, Figlio, Koplin, and Reid 1999; Rom, Peterson, and Scheve 1998; Saavedra 2000). This means that a failure to adjust benefits to compensate for inflation is equated with a decline in benefits. In contrast, Volden (2002) studies benefit competition with a logit model predicting whether a state increases its nominal benefit as a function of the proportion of neighboring states that have adopted a benefit increase.

<sup>3</sup> Below, we report evidence that our results are robust across a variety of assumptions about the value of  $D$ .

**FIGURE 1. Schematic Depiction of the Population of Concern for a Hypothetical State, S, in the Lottery Application**

Although it is theoretically feasible to use GIS to identify the location of all persons in each state's population of concern, the individual-level residential data necessary for our application are as of yet unavailable. Thus, we make the simplifying assumption that each person in a county is at the same location. This means that although the individual is the conceptual unit of analysis, our empirical analysis with GIS is conducted using counties as the units for computation.<sup>4</sup> In the following, we consider alternative assumptions about the point within a county at which its residents are located.

<sup>4</sup> Clearly, the smaller the units, the more accurate would be our calculation of the degree of concern. We chose the county because it is the lowest level of aggregation at which population counts are available annually for our period of analysis. A smaller unit, like the Census tract, is not feasible because a large portion of the country was not tracted for part of our period of analysis (i.e., for years before 1980) and boundaries for tracts change over time.

### Measuring the Propensity of Individuals to Engage in the Behavior of Concern

For each individual,  $i$ , in the population of concern of state  $s$ , we use GIS to determine the distance,  $d_i$ , of that person from the closest state that had a lottery prior to the beginning of the year. [Note that we employ the subscript  $i$  to denote a characteristic of an *individual*; subscripts  $s$  and  $c$  indicate a characteristic of a *state* or *county*, respectively.] Figure 1 depicts the value of  $d_i$  for an individual who resides at the dot near the western border of state S. Although in theory, we could measure distance with as much precision as desired using GIS, if measurement were highly precise (e.g., distance were measured to the nearest mile), there would be a dramatic increase in the memory and time required for our computer program. Thus, we measure distance to the nearest multiple of ten miles (i.e., 0 miles, 10 miles, 20 miles, etc.) [See Section I of the unpublished supplement for a detailed description of how we

use GIS to measure distance from the nearest state with a lottery.]

For each individual,  $i$ , in the population of concern, let the *propensity to play another state's lottery* be defined by

propensity to play another state's lottery,  $= f(d_i)$ ,

where  $f(d_i)$  is a decreasing monotonic function of  $d_i$  that maps a distance of 0 into a propensity score of 1 and a distance of  $D$  into a propensity of 0, and has declining slope as  $d_i$  increases, as illustrated in the top panel of Figure 2. Thus, for someone living right on the border of a state with a lottery (i.e., for whom  $d_i = 0$ ), the propensity to play another state's lottery is 1.00 (the maximum value on the propensity scale). We assume that as distance from the border rises (i.e., as  $d_i$  increases), propensity to play another state's lottery declines, but the rate of decline decreases as distance gets larger. This assumption reflects a belief that a marginal increase in distance has a greater impact on propensity when someone leaves near the state border than when someone lives farther away. For example, persons living 0 and 20 miles from a state with a lottery should have propensities to play the neighbor's lottery that differ more than persons living 80 and 100 miles from the border. When distance from the border reaches  $D$ , the propensity to play another state's lottery subsides to zero (the minimum value on the scale). Thus, although the population of concern consists of all adults living within  $D$  miles of a state with a lottery, not everyone in this population is assumed to be of equal concern to policymakers. In effect, the concern of state officials about someone going to another state to play the lottery declines as the distance of the person's residence from the nearest state with a lottery increases.

### Aggregating Individual Propensities to Compute Degree of Concern

Let the degree of concern about residents going to other states to play the lottery for a state in a year be zero if the state has its own lottery in that year. Then, for any state,  $s$ , that does not have a lottery, calculate the degree of concern by adding up the propensity to play another state's lottery over all persons in the population of concern of state  $s$ , and dividing the sum by the state's adult population:

$$\begin{aligned} \text{Degree of concern}_s &= \left[ \sum_{\text{each person, } i, \text{ in the population of concern of state } s} \text{propensity to play another state's lottery}_i \right] / (\text{adult population}_s) \\ &= \left[ \sum_{\text{each person, } i, \text{ in the population of concern of state } s} f(d_i) \right] / (\text{adult population}_s) \end{aligned}$$

So calculated, a score of 1 for degree of concern indicates the hypothetical condition in which every adult

in a state without a lottery lives right at the border of a state with a lottery; if half the adults in the state lived right at the border and the other half lived more than  $D$  miles away, degree of concern would be .50. At the other extreme, degree of concern would be zero if no adult lived less than  $D$  miles from a state with a lottery, or if the state already had its own lottery during the year of measurement.

### USING GIS TO ESTIMATE DEGREE OF CONCERN ABOUT WELFARE MIGRATION

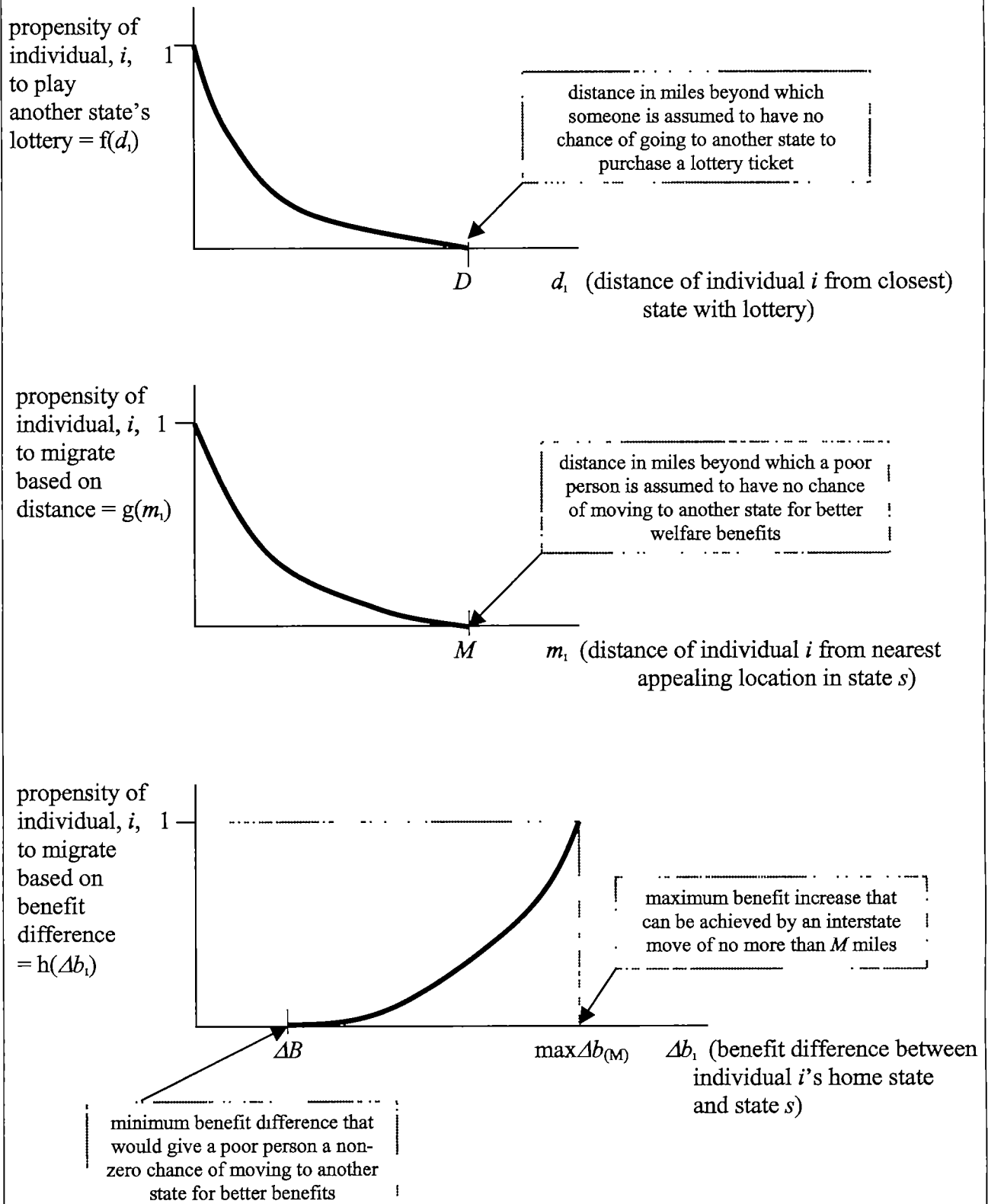
Estimating the degree of concern about welfare migration is more complex than estimating the degree of concern about residents going to other states to play the lottery. We assume that an individual's propensity to play another state's lottery is a function of just one variable: his or her distance from the nearest state with a lottery. But we assume that a poor person's propensity to move to a state for better welfare benefits is a function of two factors: the distance of the person from an appealing location in the state and the benefit increase for which the person would be eligible.

#### Identifying the Population of Concern

Let  $M$  be the maximum distance (in miles) a poor person would be willing to move for better welfare benefits. Let  $\Delta B$  be the minimum benefit difference in dollars that would motivate a poor person to move for greater welfare benefits.<sup>5</sup> Given these definitions of  $M$  and  $\Delta B$ , the population of concern for state  $s$  (those poor persons at risk of moving to state  $s$  for better welfare benefits) in a year can be defined as all poor persons in other states (neighboring or not) who live (1) within  $M$  miles of an appealing location in state  $s$ , and (2) in a state that has a welfare benefit at least  $\Delta B$  dollars lower than state  $s$ 's in the previous year.

We operationalize an "appealing location" as a city with an absolute population greater than  $p$ , and test models making varying assumptions about the value of  $p$ . One assumption is that  $p$  is zero. This implies that the poor are "location neutral"; they perceive any location as adequate if it is close enough and the welfare benefit there is sufficiently large. Most versions of the race to the bottom thesis presume that the poor are single-minded in pursuit of more generous welfare assistance (e.g., Bailey and Rom 2004; Rom, Peterson, and Scheve 1998). If this were correct, then the assumption that  $p$  equals zero would be reasonable. However, an alternative assumption is that  $p$  is substantially greater than zero, with poor persons having a clear preference for urban areas over rural ones. This is consistent with a belief that when poor persons move they choose a location that provides both generous welfare benefits and good opportunities for employment (Schram, Nitz, and Krueger 1998).

<sup>5</sup> As with  $D$  in the lottery model, our application assumes  $M$  and  $\Delta B$  are constant across individuals. But below, we report evidence of the stability of our findings across alternative assumptions about these values.

**FIGURE 2. Functions Mapping an Individual's Geographic Location into the Propensity to Engage in the Behavior of Concern for the Lottery and Welfare Models**

**FIGURE 3. Schematic Depiction of the Population of Concern for a Hypothetical State, S, in the Welfare Application**

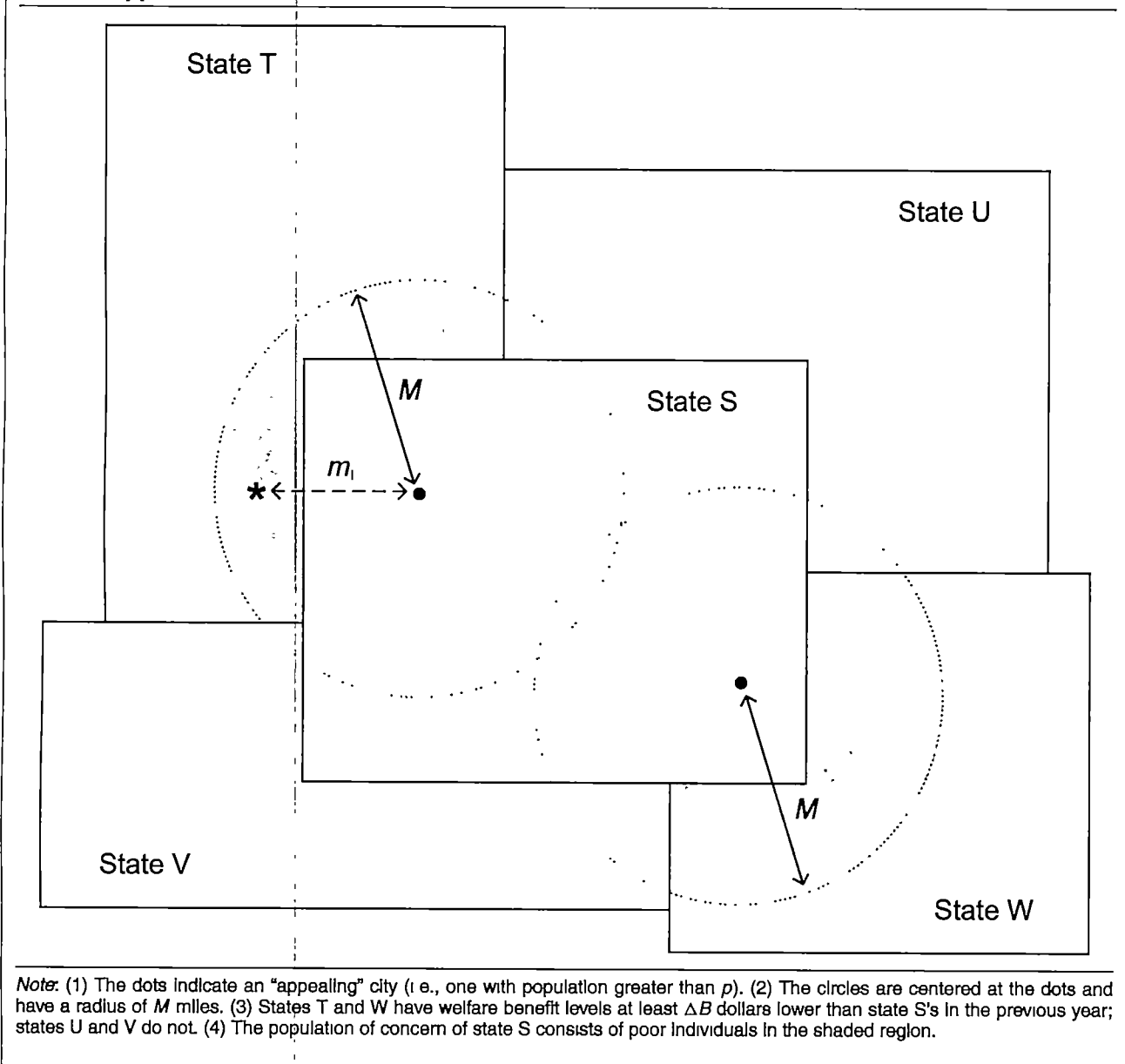


Figure 3 illustrates the population of concern for a hypothetical state, S. States T and W have welfare benefits appreciably lower (i.e., at least  $\Delta B$  dollars lower) than state S's; states U and V have benefits either higher than those of state S or lower than S's but by an amount less than  $\Delta B$  dollars. State S has two (appealing) cities with population greater than  $p$ , the locations of which are denoted by dots. The two circles are centered at these dots and have radius  $M$ . Thus, the circles represent persons living within  $M$  miles of an appealing location in state S. The shaded region represents the subset of the population within the circles that reside in states with welfare benefits appreciably (i.e., at least  $\Delta B$  dollars) lower than state S's. Therefore, poor persons residing in the shaded area constitute the population of concern of state S. Note that poor indi-

viduals in states U and V who live within the circles are not in the population of concern because benefit levels in these two states are not at least  $\Delta B$  dollars lower than the benefit in state S. As with our lottery application, individual-level residential data for conducting our welfare analysis are not available. Thus, we again assume that all persons in a county are at the same location, thereby permitting measurement of the degree of concern using county-level data.

### Measuring the Propensity of Individuals to Engage in the Behavior of Concern

For each individual in the population of concern of state  $s$ , we use GIS to determine the distance,  $m_i$ ,



of that person from the nearest appealing location in state  $s$ . [Figure 3 shows the value of  $m_i$  for an individual who resides at the asterisk in the southwest section of state  $T$ .] For each of these persons, let the *propensity to migrate based on distance* be defined by

$$\text{propensity to migrate based on distance}_i = g(m_i),$$

where  $g(m_i)$  is a decreasing monotonic function of  $m_i$  that maps a distance of 0 into a propensity score of 1 and a distance of  $M$  into a propensity of 0, as depicted in the middle panel of Figure 2. Similarly to  $f(d_i)$  in the lottery model,  $g(m_i)$  has declining slope as  $m_i$  increases (to reflect an assumption that a marginal increase in distance has a smaller impact on propensity to migrate as distance gets larger).

Next, for each person,  $i$ , in the population of concern of state  $s$ , determine the benefit difference,  $\Delta b_i$ ; this is defined as the welfare benefit in state  $s$  minus the welfare benefit in  $i$ 's home state. Let  $\max \Delta b_{(M)}$  be the maximum benefit increase (in absolute dollars) that can be achieved by an interstate move of no more than  $M$  miles during any year in the period of analysis. For each individual,  $i$ , in the population of concern, let the *propensity to migrate based on benefit difference* be defined by

$$\text{propensity to migrate based on benefit difference}_i = h(\Delta b_i),$$

where  $h(\Delta b_i)$ —depicted in the bottom panel of Figure 2—is an increasing monotonic function of  $\Delta b_i$  that maps a benefit difference of  $\Delta B$  into a propensity score of 0 and a benefit difference of  $\max \Delta b_{(M)}$  into 1, and has increasing slope as  $\Delta b_i$  increases.

In a final individual-level calculation, for each person,  $i$ , in the population of concern of state  $s$ , we determine the *propensity to migrate for better welfare benefits*:

$$\begin{aligned} \text{propensity to migrate for better welfare benefits}_i &= (\text{propensity to migrate based on distance}_i) \\ &\times (\text{propensity to migrate based on benefit difference}_i). \end{aligned}$$

Because both terms on the right side can range from zero to one, their product—the propensity to migrate for better welfare benefits—is confined to the same range. It is designed to be high only when both the propensity to migrate based on distance and the propensity to migrate based on benefit difference are high. For a poor person living (a) in a state with a benefit level  $\max \Delta b_{(M)}$  lower than state  $i$ 's and (b) zero miles from an appealing location in state  $s$ , the propensity to migrate for better welfare benefits would be 1.00 (the maximum value of the propensity scale). Someone who lives  $M$  or more miles from an appealing location in state  $s$  would have a score of zero (the minimum value on the scale) regardless of the benefit level available in his or her home state.

## Aggregating Individual Propensities to Compute Degree of Concern

If state  $s$ 's population of concern in a year is the empty set, let its degree of concern about welfare migration be zero. For other state-years, calculate the degree of concern by summing the propensity to migrate for better welfare benefits over all persons in the population of concern, and then dividing by the poor population of state  $s$ :

Degree of concern $_s$

$$\begin{aligned} &= \left[ \sum_{\text{each person, } i, \text{ in the population of concern of state } s} \text{propensity to migrate for better welfare benefits}_i \right] / (\text{poor population}_s) \\ &= \left[ \sum_{\text{each person, } i, \text{ in the population of concern of state } s} [g(m_i) \times h(\Delta b_i)] \right] / (\text{poor population}_s). \end{aligned}$$

Defined this way, degree of concern equals zero if there are no “appealing locations” in state  $s$ , if there are no poor persons in other states who live within  $M$  miles of an appealing location in state  $s$ , or if there are no nearby states with a benefit level at least  $\Delta B$  dollars lower than state  $s$ 's.

## CONSTRUCTING MULTIPLE INDICATORS OF DEGREE OF CONCERN FOR THE LOTTERY AND WELFARE ANALYSES

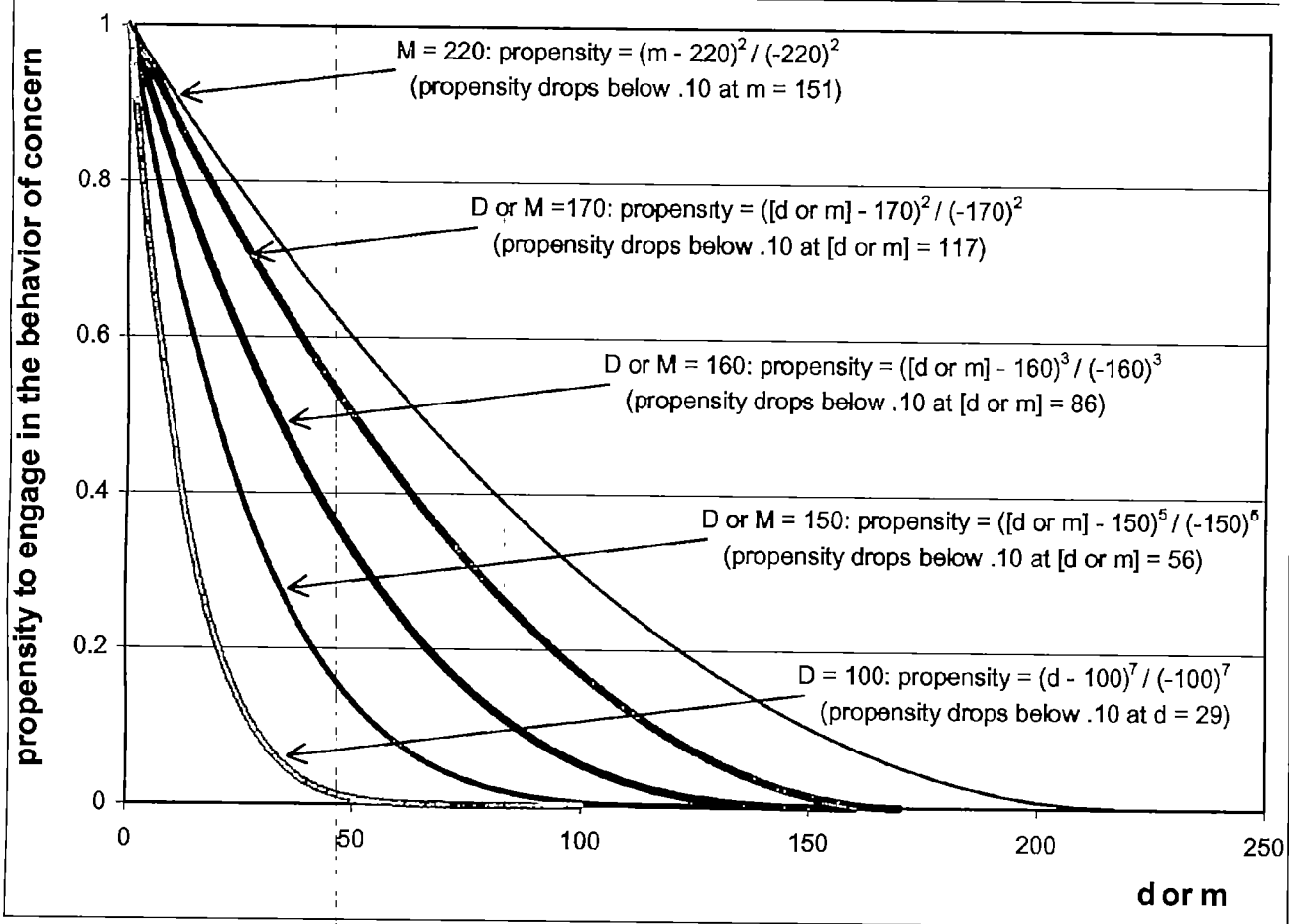
Several parameters must be assigned specific values before our degree of concern variables can be measured using GIS. For the lottery model:

- $D$ , the maximum distance (in miles) a person would be willing to travel to purchase a lottery ticket.

For the welfare model:

- $p$ , the minimum population size for a city to be appealing to the poor;
- $M$ , the maximum distance (in miles) a poor person would be willing to move for better welfare benefits; and
- $\Delta B$ , the minimum benefit difference (in dollars) that would motivate a poor person to move for better welfare benefits.

Because of theoretical uncertainty about the appropriate values for  $D$ ,  $p$ ,  $M$ , and  $\Delta B$ , we construct our indicators of degree of concern using multiple values for these parameters, so that we can test the robustness of our empirical results to assumptions that are, to some degree, arbitrary. We employ four different values of the minimum population for a city to be appealing to

**FIGURE 4. Specific Functions Mapping an Individual's Geographic Location Into Propensity to Engage in the Behavior of Concern**

the poor [ $p$ ] (0; 100,000; 200,000 and 500,000),<sup>6</sup> three different values of the minimum benefit difference required to motivate a move [ $\Delta B$ ] [100, 150, and 200 real (1995) dollars, measured using Berry, Fording, and Hanson's (2000) state cost of living index as a deflator], four different values of the maximum distance someone will travel to buy a lottery ticket [ $D$ ] (100, 150, 160, and 170 miles), and four different values of the maximum distance someone will move for better welfare benefits [ $M$ ] (150, 160, 170, and 220 miles). We calculate that  $\max \Delta b_{(M)}$ —the maximum benefit increase that can be achieved by an interstate move of no more than  $M$  miles during any year in the period of analysis—is 878 real [1995] dollars for all four values of  $M$ .

Each value of  $D$  and  $M$ , however, specifies only a maximum distance that a person is willing to travel to buy a lottery ticket or move for better welfare benefits. For each value of  $D$ , we must specify a function,  $f(d_i)$ , that maps each distance smaller than  $D$  into a value for the propensity to play another state's lottery; and for each value of  $M$  we must specify a function,  $g(m_i)$ , that maps each distance smaller than  $M$  into a value for the

propensity to migrate based on distance. Figure 4 plots these functions and presents their equations. Although three of the values for  $D$  and  $M$ —150, 160, and 170 miles—seem quite similar, the overall functions associated with these values are quite different. Indeed, the substantially different rates of descent of the functions associated with these values make it so the distance at which the propensity to engage in the behavior of concern drops below .10 varies substantially across the functions: when  $D$  or  $M$  is 150 miles, this distance is 56 miles; when  $D$  or  $M$  is 160, the distance is 86 miles; and when  $D$  or  $M$  is 170, it is 117 miles.

For each value of  $\Delta B$  (the minimum benefit difference that would motivate a move), we specify a different functional form for  $h(\Delta b_i)$ , the function that maps each benefit difference [ $\Delta b_i$ ] into a value for the propensity to migrate based on benefit difference. Each of these functions maps  $\Delta B$  (\$100, \$150, or \$200) into a score of zero, and increases monotonically, mapping  $\max \Delta b_{(M)}$  (the maximum achievable benefit increase, \$878) into a propensity value of 1. Figure S-1 in our unpublished supplement plots these functions and shows their equations.

As noted earlier, given the lack of availability of individual-level data, we assume that all persons in a county are at the same location. Therefore, to

<sup>6</sup> Because 1980 is the closest decennial Census year to the midpoint of the time periods we examine, we determine the cities surpassing the various population thresholds ( $p$ ) using 1980 Census data.

operationalize degree of concern, we must designate a specific point within a county as the assumed location of its population. Our goal is to pick a location that yields estimates of the degree of concern that are close to the values we would obtain if we were able to construct a measure taking into account the exact location of each individual. Because we cannot determine with certainty the best location, we construct measures of the degree of concern for the lottery and welfare applications assuming both (1) that all persons in a county are at its geographic center, and (2) that all persons in a county are at the center of the city with the largest population. For a county with one dominant city, a degree of concern measure based on the center of the largest city is a good choice. However, when population is more evenly dispersed, an indicator based on the geographic center of the county seems superior. Although both versions of the measure have their theoretical advantages, in practice there is little difference between the two. Across a large number of measures of degree of concern making different assumptions about the values of  $D$ ,  $p$ ,  $M$ , and  $\Delta B$  (all 52 measures for which results are reported in Tables 1 or 2, or in Tables S-1 or S-2 in our unpublished supplement), the correlation between the "geographic center" version of the measure and the "largest city" version always exceeds .98.

This result, however, does not indicate whether our measures of degree of concern yield scores similar to those that would be obtained if we were to employ the individual as the unit of analysis. To investigate this, we construct measures of the degree of concern making two different extreme assumptions about the location of population within counties. One set of measures presumes that all individuals are at the location in a county *closest* to the point of attraction to those in the population of concern (i.e., a state with a lottery, or a place that is appealing to the poor). The other set assumes that all persons are at the location in the county *farthest* from the point of attraction. Across a large number of indicators of the degree of concern, the average correlation between the "closest point" version of the measure and the "farthest point" version is .94. [Half the 52 correlations exceed .98; all but 9 are greater than .90; the lowest is .68.] Because the two maximally divergent assumptions yield measures that correlate very highly, it is very unlikely that the precise location of individuals within counties has a significant impact on the indicators we construct. This gives us confidence that the county is a unit of analysis small enough to produce measures of the degree of concern that are very likely to be close to the ones we would obtain if we knew the precise location of each individual in the population of concern. In all empirical tests that follow, we employ measures of degree of concern assuming that all persons in a county live at its geographic center.

For the lottery application, each of the four functions for  $f(d_i)$  yields a distinct measure of the propensity to play another state's lottery and ultimately a distinct value for degree of concern about residents going to other states to play the lottery. For our welfare application, using all combinations of the four values for  $p$ , the

three functions for  $h(\Delta b_i)$ , and the four functions for  $g(m_i)$  yields 48 indicators of the propensity to migrate for better welfare benefits and ultimately, the degree of concern about welfare migration.<sup>7</sup> We use ArcGIS 9.0 software (Environmental Systems Research Institute 2004) to construct our degree of concern indicators, but our calculations could be done with virtually any GIS software. We describe our GIS measurement procedure in greater detail in our unpublished supplement.

## EMPIRICAL TESTS OF THE LOTTERY LEARNING AND COMPETITION HYPOTHESES

We test the lottery learning hypothesis by replicating the estimation of B&B's (1990) discrete event history model, which has as dependent variable whether a state adopts a lottery in a year, and includes among the independent variables the number of previously adopting neighbors. In particular, we estimate the model B&B report in the right-most column of their Table 1 (p. 406). To test the lottery competition hypothesis, we substitute each of our four measures of degree of concern for the number of previously adopting neighbors. To allow for duration dependence (i.e., to permit the probability of adoption of a lottery to vary over time), we also add a time counter to each equation (Buckley and Westmoreland 2004).<sup>8</sup> We estimate each equation with probit, using the same 901 annual observations of the 48 contiguous states for 1964 through 1986 employed by B&B, but we compute robust standard errors—clustering by state to allow for dependence of observations within states.

The learning hypothesis receives empirical support, but due to the addition of the time counter, the estimated effect of number of previously adopting neighbors is somewhat smaller than reported in B&B's article. The coefficient for number of previously adopting neighbors (MLE = 0.135), although failing a test of statistical significance at .05, is significantly positive at the .10 level (one-tailed test;  $Z = 1.29$ ).

<sup>7</sup> The computation of degree of concern for our lottery analysis requires county counts of adult population (for each year during the period of analysis, 1964–86). Using population age 18 or older would be ideal because the vast majority of states limit sales of lottery tickets to this age group, but data constraints mandate that we employ population twenty or older. The calculation of degree of concern for our welfare analysis requires county counts of poor population (for each year during the period 1961–90). We obtained 1960 Census adult population data (Haines 2005) and annual estimates for the period 1970–90 (U.S. Bureau of the Census 2004); values for years between 1960 and 1970 were calculated using linear interpolation. Poverty counts for Census years 1960, 1970, 1980 and 1990 were obtained from Rural Policy Research Institute (2003), and values for intercensal years were estimated via linear interpolation between Census values.

<sup>8</sup> This time counter is 1 in 1964 (the first year of analysis), 2 in 1965, ..., and 23 in 1986 (the last year of analysis). Note that we also estimated equations including dummy variables for years (as suggested by Beck, Katz, and Tucker [1998]). The estimated effects of degree of concern and number of previously adopting neighbors are very similar regardless of which approach to allow for duration dependence is used; thus, we report results for the more parsimonious specifications with a single time count variable.

Results regarding the competition hypothesis (i.e., coefficient estimates for degree of concern when it is substituted for number of previously adopting neighbors) are summarized in columns 2 through 4 of Table 1. The maximum likelihood estimates of the coefficients for degree of concern (see column 2) are positive as predicted, and each is significant at the .05 level.<sup>9</sup> When the degree of concern about residents going to other states to play the lottery increases from its fifth percentile value in the sample to its ninety-fifth percentile value (and all other independent variables in the model are held constant at the mean), the probability that the state will adopt a lottery in a given year is estimated to increase by somewhere from .021 to .027 (depending on the value assumed for  $D$ ; see column 4).<sup>10</sup> Although such probability differences may seem small, they are appreciable, given how rare lottery adoptions are during our period of analysis: the estimated probability that a state without a lottery will adopt one in a year is .030 (27 adoptions out of 901 observed cases). Thus, empirical analysis of lottery adoptions is consistent with both the competition and learning hypotheses, but support for the competition hypothesis is somewhat stronger. Yet because number of previously adopting neighbors and degree of concern are highly correlated—between .72 and .86, depending on which of the four versions of the latter variable is used—models in which just one of these variables is included may indicate an effect of that variable when none is present because the variable is picking up the true effect of the excluded variable.

A more definitive test of the two hypotheses can be achieved by including both degree of concern and number of previously adopting neighbors in the same model. If diffusion occurs strictly because of learning (and interstate competition plays no role whatsoever), the relationship between number of previously adopting neighbors and the probability of an adoption should survive a control for degree of concern, but the relationship between degree of concern and adoption probability should decline to near zero when number of previously adopting neighbors is added to the model. In contrast, if diffusion were solely a result of competition (and learning were not a factor), the relationship between degree of concern and the probability of adoption should survive a control for number of previously adopting neighbors, but the relationship between the neighbors variable and the probability of adoption should disappear when controlled for the effect of degree of concern.

Columns 5 through 9 of Table 1 present results when we include both number of previously adopting neighbors and degree of concern in our models. Observe that with degree of concern included, the coefficient estimate for number of previously adopting neighbors declines to near zero (indeed becomes slightly negative) in all four versions of the model (see column 5). However, we can see that the estimates of the coefficients for degree of concern survive a control for number of previously adopting neighbors: indeed the MLEs for degree of concern increase somewhat when number of neighbors is added to the model (compare columns 7 and 2). Similarly, the estimates of the effect of degree of concern on the *probability* of adoption increase for all four values of  $D$  when number of previously adopting neighbors is added (compare columns 9 and 4). Although point estimates of the effect of degree of concern increase in magnitude when number of previously adopting neighbors is included,  $Z$  statistics for the MLEs for degree of concern decrease (compare columns 8 and 3) and the width of confidence intervals for predicted changes in the probability of an adoption widen (compare columns 9 and 4). This is likely due to the strong correlation between number of previously adopting neighbors and degree of concern. Thus, there is compelling evidence that the interstate influence leading to the diffusion of the lottery results from competition—fear by state officials of losing revenues to neighboring states—rather than policy learning.<sup>11</sup>

## EMPIRICAL TESTS OF THE WELFARE LEARNING AND COMPETITION HYPOTHESES

We test the welfare learning hypothesis by replicating the estimation of BFH's original model, which has as dependent variable a state's AFDC benefit in a year (i.e., its maximum monthly AFDC payment [in real dollars] for a family of four with no income), and includes among the independent variables a state's benefit relative to the average benefit in neighboring states in the previous year (for short, "benefit relative to neighbors"). In particular, we reestimate the version of the model reported in the top panel of BFH's (2003) Table 1, using the same 1,440 pooled annual observations of the 48 continental states for the 1961–90 period. BFH report 2SLS estimates for a two-equation model, treating the poverty rate as a second endogenous variable. We use OLS regression for our single-equation model. Note, however, that BFH's results change only slightly

<sup>9</sup> Although the fit of our lottery model is fairly similar across analyses assuming different functions,  $f(d)$ , mapping distance from the closest state with a lottery to the propensity to play another state's lottery, if fit had been substantially better using one of these functions, this might have provided insight into state officials' perceptions of how far people are willing to travel to play the lottery.

<sup>10</sup> This predicted increase in the probability of adoption is also statistically significant in two of the four versions of the model, as the 95% confidence interval for the estimated change in probability excludes zero for these models. For the other two versions, the 95% confidence interval just barely extends into the negative range (no further than  $-.001$ ).

<sup>11</sup> Additional support for the claim that the diffusion of the lottery is due to competition rather than learning is derived from a statistical test developed by Davidson and MacKinnon (1993) and implemented by Greene (1995, 422). For all four values of  $D$ , concerning of the equation including number of previously adopting neighbors and the equation including degree of concern as rival models, when the model involving number of neighbors is treated as the null, it is easily rejected in favor of the model including degree of concern (with  $t$ -ratios ranging in magnitude from 2.33 to 3.13). However, when the model including degree of concern serves as the null, it cannot be rejected for any value of  $D$  ( $t$  is always less than 0.22).

TABLE 1. Probit Results for Testing the Lottery Learning and Competition Hypotheses

(1) Value Assumed for <i>D</i>	(2) MLE for Degree of Concern	(3) Z Statistic for Degree of Concern <sup>a</sup>	(4) Models with Degree of Concern Substituted for Number of Previously Adopting Neighbors	(5) MLE for Number of Previously Adopting Neighbors	(6) Z Statistic for Number of Previously Adopting Neighbors <sup>a</sup>	(7) MLE for Degree of Concern	(8) Z Statistic for Degree of Concern <sup>a</sup>	(9) Change In Probability of Adoption Associated with Change in Degree of Concern from 5th to 95th Percentile (when other variables held constant at mean) <sup>b</sup>
100	1.601*	3.00	Change in Probability of Adoption Associated with Change in Degree of Concern from 5th to 95th Percentile (when other variables held constant at mean) <sup>b</sup> <b>0.021</b> [.003, .064]	-0.036	-0.28	1.744*	2.59	<b>0.023</b> [.003, .067]
150	1.081*	2.12	<b>0.022</b> [.000, .078]	-0.093	-0.60	1.427*	1.92	<b>0.031</b> [-.001, .113]
160	0.822*	1.83	<b>0.027</b> [-.001, .095]	-0.112	-0.65	1.193	1.62	<b>0.048</b> [-.004, .185]
170	0.689*	1.73	<b>0.025</b> [-.000, .086]	-0.112	-0.64	1.023	1.52	<b>0.048</b> [-.007, .178]

Note. Results are obtained using the probit procedure (with the cluster option) in Stata 8 (Tomz, Wittenberg, and King 2003). This table reports coefficient estimates for only those variables that are relevant for testing the lottery learning and competition hypotheses. The other variables assumed to influence the probability that a state will adopt a lottery are the fiscal health of the state, the proximity to gubernatorial elections, per capita income, and the proportion of the population adhering to fundamentalist religions. A full set of coefficient estimates for each equation is reported in our unpublished supplement.

<sup>a</sup> Z statistics are based on robust standard errors, clustering by state.

<sup>b</sup> The values in bold are point estimates of the change in probability, they are followed by a 95% confidence interval as calculated by Clarify 2 with Stata 8

\*  $p < .05$  (one-tailed).

**TABLE 2. Regression Results for Testing the Welfare Learning and Competition Hypotheses**

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Model with Degree of Concern Substituted for Benefit Relative to Neighbors		Model Including Both Benefit Relative to Neighbors and Degree of Concern			
Value Assumed for <i>M</i>	Slope Estimate for Degree of Concern <sup>a</sup>	<i>t</i> -ratio for Degree of Concern	Slope Estimate for Degree of Concern <sup>a</sup>	<i>t</i> -ratio for Degree of Concern	Slope Estimate for Benefit Relative to Neighbors	<i>t</i> -ratio for Benefit Relative to Neighbors
Equations Assuming $p = 0$ :						
150	-0.087	-0.81	0.006	0.05	-39.74*	-2.33
160	-0.105	-1.50	-0.048	-0.70	-36.41*	-2.13
170	-0.100*	-1.97	-0.062	-1.24	-33.71*	-1.96
220	-0.092*	-2.35	-0.065*	-1.68	-31.44*	-1.84
Equations Assuming $p = 100,000$ :						
150	-0.088	-0.24	0.091	0.26	-40.15*	-2.36
160	-0.089	-0.40	0.042	0.19	-40.15*	-2.36
170	-0.086	-0.59	0.004	0.03	-39.64*	-2.33
220	-0.092	-0.90	-0.028	-0.28	-38.37*	-2.25
Equations Assuming $p = 200,000$ :						
150	-2.739*	-1.70	-1.971	-1.22	-35.65*	-2.08
160	-0.498	-0.85	-0.188	-0.32	-38.30*	-2.22
170	-0.193	-0.69	-0.041	-0.14	-38.95*	-2.26
220	-0.135	-0.76	-0.033	-0.18	-38.74*	-2.25
Equations Assuming $p = 500,000$ :						
150	-4.399*	-2.53	-3.801*	-2.20	-35.62*	-2.09
160	-1.200*	-1.72	-0.924	-1.33	-35.92*	-2.09
170	-0.519	-1.57	-0.381	-1.16	-36.12*	-2.10
220	-0.364*	-1.73	-0.269	-1.28	-35.33*	-2.05

*Note:* All equations are estimated using the same 1,440 observations employed by BFH (2003) and assuming that  $\Delta B = 150$ . Results are obtained using the *xtpcse* procedure in Stata 8. *t*-ratios are computed using panel-corrected standard errors. This table reports coefficient estimates for only those variables that are relevant for testing the welfare learning and competition hypotheses. Among the other variables assumed to influence a state's welfare benefit are the state's poverty rate, its wage levels, several variables reflecting economic conditions in the state and in neighboring states, whether the state has a residency requirement, a set of dummy variables for the states, and a lagged dependent variable. A full set of coefficient estimates for each equation is reported in our unpublished supplement.

<sup>a</sup> Coefficients in this column are in 1,000s.

\* $p < .05$  (one-tailed).

when their welfare benefit equation is estimated using OLS.

To test the welfare competition hypothesis, we substitute for benefit relative to neighbors each of our 48 measures (relying on varying assumptions about the values of  $M$ ,  $\Delta B$ , and  $p$ ) of the degree of concern by state officials about welfare migration. Because findings tend to be stable across the three values of  $\Delta B$  (\$100, \$150, and \$200), to save space, we report results only for models assuming  $\Delta B = 150$ .

The learning hypothesis receives empirical support. The estimated coefficient for benefit relative to neighbors is -39.52, a value statistically significant in the predicted direction at the .05 level ( $t = -2.30$ ). This estimate implies that when a state's AFDC benefit relative to its neighbors increases from its fifth percentile value in the sample to its ninety-fifth percentile value (and all other independent variables in the model are held constant), the state is expected to reduce its monthly benefit for a family of four by about 34 (1995) dollars in the first year.<sup>12</sup>

BFH's (2003) benefit relative to neighbors variable is based on a weighted average (by population) of the benefit level in neighboring states. This is appropriate for a welfare learning model that assumes that states pay more attention to the benefit level in a populous neighbor than in a state of smaller size. A different learning model would assume that states are equally attentive to each of their neighbors, viewing all their neighbors as reasonable "benchmarks" regardless of their size. To test this alternative learning model, we estimate an equation that employs a measure of benefit relative to neighbors based on the *unweighted* average of benefits in contiguous states. For this equation too, the coefficient estimate for benefit relative to neighbors is negative and statistically significant.

The results regarding the welfare competition hypothesis—in which concern about welfare migration is substituted for benefit relative to neighbors—are presented in columns 2 and 3 of Table 2. The coefficient estimates for degree of concern are uniformly negative as predicted, and 6 of the 12 are statistically significant

<sup>12</sup> The mean monthly AFDC benefit for a family of four across state-years in our sample is \$660. Because there is a lagged dependent vari-

able in the model, estimated coefficients for independent variables reflect immediate impacts, the total effects of which are distributed over time (Gujarati 1995, 599–600).

at the .05 level. Yet, even the coefficient estimates that are statistically significant do not indicate that states respond to increases in concern about welfare migration with substantial decreases in welfare benefits. We calculate the predicted reduction in welfare benefits resulting from a change in the degree of concern from its fifth percentile value in the sample to its ninety-fifth, when all other independent variables are held constant, based on each of the six regressions in which the coefficient estimate for degree of concern is statistically significant. The predicted response ranges from a decrease of about four (1995) dollars in the monthly AFDC benefit for a family of four in the first year to a decrease of around \$13—with the average response being slightly more than \$7. This average decrease of \$7 is about one fifth the size of the predicted \$34 decrease resulting from an increase in benefit relative to neighbors from the fifth percentile to the ninety-fifth. Therefore, in the case of welfare, our empirical analysis offers stronger evidence for the learning hypothesis than the competition hypothesis.

However, a more compelling test of the two hypotheses can be accomplished by including both degree of concern and benefit relative to neighbors in the same model. Column 6 of Table 2 shows coefficient estimates for benefit relative to neighbors in such models. When degree of concern is controlled, all 16 slope coefficient estimates for benefit relative to neighbors attain statistical significance, and the average change in their magnitudes is a decline of just 6%. In contrast, of the six models in which the estimated effect of degree of concern is statistically significant when benefit relative to neighbors is not in the model, only two show degree of concern still significant after benefit relative to neighbors is added (see column 4); and across the six models, the coefficient estimate for degree of concern decreases in magnitude, on average, by 26%. Thus, there is little empirical evidence suggesting that states compete vigorously over benefit levels, adjusting benefits substantially in response to a concern about the potential migration of the poor. However, our data analysis strongly supports the policy-learning hypothesis that policymakers view welfare benefits in neighboring states as benchmarks for determining a reasonable benefit level and adjust their benefit levels accordingly.

In 1996, Congress eliminated AFDC and replaced it with Temporary Assistance for Needy Families (TANF). We see no reason for expecting state officials to be more or less prone to search for benchmarks for their welfare benefits under TANF than under AFDC, and thus no reason to believe that welfare policy learning has changed with the introduction of TANF. However, the incentives for state officials to compete over welfare benefits have changed significantly since the adoption of TANF. Although states set their own benefit levels under AFDC, and they continue to do so under TANF, the categorical grant for AFDC made it so that the federal government paid for at least half of every dollar of benefit provided to each AFDC recipient. In contrast, the block grant for TANF makes it so that the marginal cost of an increase in welfare benefits is borne completely by the states,

and the marginal benefit of a benefit decrease accrues exclusively to the states. Thus, one might reasonably expect that the incentive for state officials to compete over welfare benefits has increased. On the other hand, under the tight federal controls of AFDC, any competition among states was largely confined to the benefit levels set. TANF affords states much greater autonomy over a range of provisions of their welfare programs, including the discretion to set time limits, impose personal responsibility requirements and create work incentives. Because under TANF, states can discourage welfare migration by adjusting other provisions of their programs, this might lessen the motivation to undercut other states' benefits. In effect, under TANF, the focus of competition may have shifted from benefit levels to other dimensions. Thus, it is not obvious whether our evidence of a lack of vigorous benefit competition would extend to the contemporary welfare environment. On this issue, we must await empirical tests based on data from the TANF era.

## CONCLUSION

The diffusion of public policy across the American states detected in a wide variety of policy areas has been attributed to two fundamentally different processes: policy learning and economic competition. Most empirical analyses of interstate influence have used a simple specification that assumes that states are influenced equally by all their neighbors, and uninfluenced by states that are not contiguous. Such a specification is reasonable for some versions of a learning model, but is generally inappropriate for testing a competition model because when a policy diffuses due to competition, states should be influenced by other states to varying degrees depending on the size and location of specific populations within the states. We have shown how geographic information systems can be used to test competition models of policy diffusion.

Although previous studies have claimed that the diffusion of both lottery adoptions and welfare benefit levels is due to interstate competition, using GIS we find evidence of competition only in the case of the lottery. We cannot say definitively why this is so, but we can speculate. Both our lottery and welfare models assume that for many individuals, the propensity to engage in the behavior of concern is appreciably greater than zero. In the case of the lottery, there is little doubt that this assumption is true. It is clear not only that state officials perceive that people living near a state border will cross it to play the lottery, but that many people do travel to other states for this purpose (Fink and Rork 2003, Mikesell 1987). It is far less certain that a significant number of poor persons migrate for better welfare benefits. Critics have challenged the hypothesis that the poor migrate for more generous public assistance on a variety of grounds—arguing that it ignores the substantial costs of relocation (which are especially burdensome on individuals with few economic resources), and assumes unreasonably that the poor are more concerned with welfare benefits than with private economic opportunities (Schram, Nitz,



and Krueger 1998). Indeed, the most common result of individual-level research is that there is relatively little such migration (e.g., Allard and Danziger 2000; for a review of this literature, see Allard 1998).<sup>13</sup> Although it is possible that state officials believe that substantial welfare migration occurs even if it does not, and set policy based on this belief, our empirical evidence is consistent with an argument that policymakers do, in fact, recognize that large-scale migration of the poor is not a realistic concern.

The GIS techniques we have employed to study state policymaking would also be useful to political scientists and policy analysts studying competition among other jurisdictions. Indeed, there is a large literature on competition among local governments (e.g., Peterson 1981, Schneider 1989). One application of our approach would be a study of whether municipalities compete to have low sales tax rates due to a concern about residents going to a nearby jurisdiction to shop. An international relations application might assess whether immigration policies diffuse among European Union (EU) nations due to concern about individuals in nearby less developed non-EU countries immigrating for better economic opportunities. With GIS, the possibilities for careful empirical tests of models of spatial influence are vast.

## REFERENCES

- Allard, Scott W. 1998. "Revisiting *Shapiro* Welfare Magnets and State Residency Requirements in the 1990s" *Publius* 28 (Summer): 45–65.
- Allard, Scott W., and Sheldon Danziger. 2000. "Welfare Magnets: Myth or Reality?" *Journal of Politics* 62 (May): 350–68.
- Bailey, Michael A. 2005. "Welfare and the Multifaceted Decision to Move" *American Political Science Review* 99 (February): 125–35.
- Bailey, Michael A., and Mark Carl Rom. 2004. "A Wider Race? Interstate Competition across Health and Welfare Programs" *Journal of Politics* 66 (May): 326–47.
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42 (October): 1260–88.
- Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84 (June): 395–415.
- Berry, William D., Richard C. Fording, and Russell L. Hanson. 2000. "An Annual Cost of Living Index for the American States, 1960–95." *Journal of Politics* 62 (May): 550–67.
- Berry, William D., Richard C. Fording, and Russell L. Hanson. 2003. "Reassessing the 'Race to the Bottom' in State Welfare Policy. Resolving the Conflict Between Individual-Level and Aggregate Research." *Journal of Politics* 65 (May): 327–49.
- Boehmke, Frederick J., and Richard Witmer. 2004. "Disentangling Diffusion: The Effects of Social Learning and Economic Competition on State Policy Innovation and Expansion." *Political Research Quarterly* 57 (March): 39–51.
- Buckley, Jack, and Chad Westerland. 2004. "Duration Dependence, Functional Form, and Correct Standard Errors: Improving EHA Models of State Policy Diffusion." *State Politics and Policy Quarterly* 4 (Spring): 94–114.
- Cho, Wendy K. Tam. 2003. "Contagion Effects and Ethnic Contribution Networks." *American Journal of Political Science* 47 (April): 368–87.
- Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Environmental Systems Research Institute. 2004. ArcGIS 9.0. Redlands, CA: ESRI.
- Figlio, David N., Van W. Kolpin, and William E. Reid. 1999. "Do States Play Welfare Games?" *Journal of Urban Economics* 46 (November): 437–54.
- Fink, Stephen, and Jonathan Rork. 2003. "The Importance of Self-Selection in Casino Cannibalization of State Lotteries." *Economics Bulletin* 8 (10): 1–8.
- Gimpel, James G., and Jason E. Schuknecht. 2003. "Political Participation and the Accessibility of the Ballot Box" *Political Geography* 22 (June): 471–88.
- Glick, Henry R., and Scott P. Hays. 1991. "Innovation and Reinvention in State Policymaking: Theory and the Evolution of Living Will Laws" *Journal of Politics* 53 (August): 835–49.
- Greene, William H. 1995. *LIMDEP Version 7.0 User's Manual*. Bellport, NY: Econometric Software, Inc.
- Grossback, Lawrence J., Sean Nicholson-Crotty, and David A. M. Peterson. 2004. "Ideology and Learning in Policy Diffusion" *American Politics Research* 32 (September): 521–45.
- Gujarati, Damodar N. 1995. *Basic Econometrics*. 3rd ed. New York: McGraw-Hill.
- Haines, Michael R., and the Inter-university Consortium for Political and Social Research. 2005. Historical, Demographic, Economic, and Social Data: The United States 1790–2000 [Computer file] (Study #2896). ICPSR version. Hamilton, NY: Colgate University/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producers], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005.
- Ka, Sangjoon, and Paul Teske. 2002. "Ideology and Professionalism. Electricity Regulation and Deregulation Over Time in the American States." *American Politics Research* 30 (May): 323–43.
- Mikesell, John L. 1987. "The Effect of Maturity and Competition on State Lottery Markets." *Journal of Policy Analysis and Management* 6 (Winter): 251–53.
- Mooney, Christopher Z., and Mei-Hsien Lee. 1995. "Legislative Morality in the American States: The Case of Pre-Roe Abortion Regulation Reform." *American Journal of Political Science* 39 (August): 599–627.
- Peterson, Paul E. 1981. *City Limits*. Chicago: University of Chicago Press.
- Rom, Mark Carl, Paul E. Peterson, and Kenneth F. Scheve, Jr. 1998. "Interstate Competition and Welfare Policy." *Publius* 28 (Summer): 17–37.
- Rural Policy Research Institute. 2003. "Poverty Rates 1960–2000." <http://csrc.rupri.org/animation/povertyrate.html> (April 3, 2003).
- Saavedra, Luz Amparo. 2000. "A Model of Welfare Competition with Evidence from AFDC." *Journal of Urban Economics* 47 (March): 248–79.
- Schneider, Mark. 1989. *The Competitive City. The Political Economy of Suburbia*. Pittsburgh: University of Pittsburgh Press.
- Schram, Sanford F., Lawrence Nitz, and Gary Krueger. 1998. "Without Cause or Effect: Reconsidering Welfare Migration as a Policy Problem." *American Journal of Political Science* 42 (January): 210–30.
- Simon, Herbert A. 1997. *Administrative Behavior*. 4th ed. New York: Free Press.
- Tomz, Michael, Jason Wittenberg, and Gary King. 2003. "CLARIFY: Software for Interpreting and Presenting Statistical Results." Version 2.1. Stanford University, University of Wisconsin, and Harvard University <http://gking.harvard.edu> (January 5, 2003).
- Tweedie, Jack. 1994. "Resources Rather than Needs: A State-Centered Model of Welfare Policy Making." *American Journal of Political Science* 38 (August): 651–72.
- United States Bureau of the Census. 2004. "U.S. Census Bureau, Population Estimates, Archives." <http://www.census.gov/popest/archives> (August 13, 2004).
- Volden, Craig. 2002. "The Politics of Competitive Federalism: A Race to the Bottom in Welfare Benefits?" *American Journal of Political Science* 46 (June): 352–63.
- Walker, Jack L. 1969. "The Diffusion of Innovations Among the American States" *American Political Science Review* 63 (September): 880–99.

<sup>13</sup> Even Bailey (2005, 134), who claims that this research underestimates the amount of welfare migration, concludes that although "welfare benefits exert a nontrivial effect on state residential choice," the poor do not move in large numbers for greater welfare benefits.

## Ethnicity and Trust: Evidence from Russia

DONNA BAHRY *The Pennsylvania State University*

MIKHAIL KOSOLAPOV *Institute of Sociology, Russian Academy of Sciences*

POLINA KOZYREVA *Institute of Sociology, Russian Academy of Sciences*

RICK K. WILSON *Rice University*

*The willingness to trust strangers has been associated with a variety of public benefits, from greater civic-mindedness and more honest government to higher rates of economic growth, and more. But a growing body of research finds that such generalized trust is far more common in ethnically homogeneous than in more diverse societies. Ethnic difference is believed to breed more particularistic, ingroup ties, thus undermining both generalized and cross-ethnic trust. We argue that this image is too narrow, and we propose a broader model to identify the factors that give rise to cross-ethnic trust. Using data from two minority regions of Russia, we find considerable support for the model. We also find that high ingroup or particularistic trust is no barrier to faith in another ethnic group.*

Trust has increasingly come to be recognized as a critical element of both democracy and markets. The willingness to trust strangers promotes civic engagement and community building, and helps overcome the dilemmas of collective action (Fukuyama 1995; Putnam 1993; Uslaner, 2002). It also plays a central role in economic life, fostering cooperation and thus facilitating impersonal exchange. The results can be dramatic: higher trust has been associated with greater citizen involvement in politics, lower corruption, more effective public services, higher economic growth, and other benefits (see, e.g., Knack and Keefer 1997; LaPorta et al. 1997; Zak and Knack 2001).

But such generalized faith in others seems to be far greater in ethnically homogeneous than in more diverse societies. Cross-national surveys demonstrate that trust is lower in heterogeneous countries (Knack and Keefer 1997). Research in the United States points to less generalized faith in others when local communities are diverse (Alesina and La Ferrara 2002). Studies of other elements of social capital come to a similar conclusion (Alesina and La Ferrara 2000; Costa and Kahn 2003).

The dominant explanation is that ethnic difference breeds more particularized, rather than generalized, trust. To use Fukuyama's (1999) phrase, the radius of particularized trust is short. People extend their confidence to a narrow set of ingroups—to family, friends,

and others like themselves, but seldom beyond. Ethnic difference is thus assumed to generate a high level of ingroup trust, but little or no confidence in others. Some research suggests that the relationship is actually zero-sum: the higher the trust in one's own group, the lower the faith in people outside it.

This image of exclusionary ethnic trust is problematic, however. Because most people are assumed to have faith in members of their own ethnic group, cross-ethnic confidence must be low by definition. If such trust does arise, the assumption is that faith in the ingroup cannot be high. The possibility of what might be called "inclusionary" trust—in both one's own and in other groups—is omitted. So, too, is the possibility of atomization, or distrust of in- and outgroups alike. Also missing is the very real possibility that cross-ethnic trust might be selective; some outgroups may be viewed more favorably than others.

Thus, a fuller model is essential if we are to explain when and where cross-ethnic confidence *does* arise. We think the first step is to reevaluate the connections among ethnicity and generalized versus particularized trust. We then offer a model to identify the micro-foundations of confidence across ethnic lines. We draw on new survey data from two multiethnic republics of Russia, to determine what factors lead culturally and racially diverse groups to trust others. Given the spread of ethnic assertiveness in the USSR and its successor states over the past two decades, this should provide a particularly stringent test of our model.

The analysis can also offer some purchase on how diversity affects social and economic outcomes more broadly. Research on cross-national variation in social and economic performance has tied ethnic difference to lower rates of economic growth, lower provision of public goods, higher corruption, and other economic problems (see, e.g., Alesina, Baqir, and Easterly 1999; Easterly and Levine 1997; Zak and Knack 2001). The logic in most of the literature is similar to that for trust: ethnic difference is seen as impeding cooperation. In Paul Collier's (2001, 130) words, "The underlying propositions are that ethnic divisions make cooperation more difficult and victimization more likely." The results might be interpreted as a brief for separation.

Donna Bahry is Professor and Head, Department of Political Science, The Pennsylvania State University, 219 Pond Lab, University Park, PA 16801 (DBahry@PSU.edu).

Mikhail Kosolapov is Senior Scientist, Institute of Sociology, Russian Academy of Sciences, Krzhizhanovskogo 24/35 k. 1, Moscow, Russia.

Polina Kozyreva is Senior Scientist, Institute of Sociology, Russian Academy of Sciences, Krzhizhanovskogo, 24/35 k. 1, Moscow, Russia.

Rick Wilson is Professor and Chair, Department of Political Science, Rice University, Houston, TX (RKW@Rice.edu).

Data for this paper were collected with the support of the National Science Foundation (OPP0082715) and the National Council for Eurasian and East European Research. We would also like to thank the Rockefeller Foundation's Bellagio Center and the Russell Sage Foundation for their support of the analysis, and Jim Gibson and three anonymous reviewers for comments on an earlier draft.

But as with interethnic trust, the literature is less clear on when and where ethnic groups do cooperate. Thus, identifying the factors that promote or impede cross-ethnic confidence can also help us understand other aspects of cooperation across ethnic lines.

## GENERALIZED VERSUS PARTICULARIZED TRUST

A growing body of research on trust emphasizes the distinction between generalized versus particularized confidence in others—part of what Putnam (2000) labels “bridging” versus “bonding” social capital. Generalized trust helps promote the norms of reciprocity and cooperation that underpin civil society (Putnam). It appears to reflect an individual’s belief that most others share the same fundamental values, and belong to the same “moral community” (Fukuyama 1995; Uslaner 2002).

Particularized trust entails deeper ties to a closer circle such as family members, friends, and others with similar backgrounds. Particularized trusters, as Uslaner (2002) argues, tend to be suspicious of people they don’t know; and feel they have little control over what happens. They are also more withdrawn from society at large (Uslaner; Uslaner and Conley 2003).

The difference is considered to be crucial for community building and public decision making. Generalized trusters appear to engage more readily in the community and in collective action, and cooperate more easily with people from different backgrounds. Generalized faith in others also seems to have a far more positive impact on the spread of information and innovation. The broader an individual’s connections, the more access to new ideas (Granovetter 1973).

But a number of authors have argued that generalized trust is diminished in ethnically diverse societies.<sup>1</sup> The assumption, as Alesina and La Ferrara (2002, 207) note, is that “most individuals are less inclined to trust those who are different from themselves.” Ethnic difference is described as breeding ingroup rather than outgroup confidence, due either to groups’ different worldviews or preferences; greater ability to coordinate within groups rather than across them; or simply aversion toward outgroup members (see, e.g., Easterly and Levine 1997; Knack and Keefer 1997; Landa 1995).<sup>2</sup> The implication, then, is that people who “bond” are less likely to “bridge.” Intraethnic trust is assumed to be the inverse of confidence in others at large and confidence in outgroups.

We think the image of a tradeoff is persuasive but incomplete. Ingroup trust need not be an impediment to confidence in outgroups or in others generally (cf. Herring, Jankowski, and Brown, 1999). Some people may trust only their own and distrust outsiders, but others may well trust both. In fact, those with

generalized faith in others should be disposed to such inclusionary trust.<sup>3</sup>

The tradeoff thesis poses several other problems as well. One is the assumption of ingroup homogeneity: groups are in conflict because they have cohesive and opposing preferences. But cohesion clearly varies. In fact, as Hardin (1995) observes, ingroup norms—the informal rules that reinforce a sense of distinctiveness—arise as a way of reducing differences in degrees of commitment within the group.

The idea of a tradeoff in trust also implies that ingroup attachment breeds a similar reaction to all outgroups. In part, it does: what Hardin labels “norms of exclusion” would shape a person’s trust toward outgroups in general. But we also know that attitudes have a selective element as well. People may respond to various groups differently—depending in part on images of the other group (Fiske 2000, 2002) and on everyday contact (see Gibson 2004; Oliver and Wong 2003; Welch et al. 2001, for examples).

Disentangling these issues has been difficult, however, because individual-level data on cross-ethnic trust are rare.<sup>4</sup> Research linking ethnic difference to confidence in others focuses primarily on generalized trust (see, e.g., Alesina and La Ferrara 2002; Knack and Keefer 1997). Most such studies find low generalized trust and conclude from it that confidence across ethnic lines must be low as well. But this can be misleading, because people may display more trust in an outgroup than in “other people” as a whole.

Thus, explaining cross-ethnic trust requires that we measure it directly, rather than infer it from questions about generalized faith in others. That would allow us to determine whether and how the two forms of confidence are in fact related. We would also need a model to help distinguish between the disposition to trust outgroups as a whole versus the disposition to trust a specific outgroup.

We think that such a model should include three components. The first would be the disposition to broad-gauged trust, without reference to ethnicity. The second would be the disposition to trust other ethnic groups, based on one’s own ethnic attachment. And the third would be the disposition to trust a particular outgroup. Each of these, in turn, includes several elements.

Broad-gauged trust would include generalized faith in others, conceived here as a core value that persists over time; localized or “intermediate trust,” focused in the community at large; and confidence in government. All of these may be related empirically, but they

<sup>1</sup> There is debate, however, over which ethnic “structures” (numbers and size distribution of ethnic groups) are most problematic. See, for example, the discussion in Collier (2001); Fearon (2003); and Garcia-Montalvo and Reynal-Querol (2002).

<sup>2</sup> For an overview, see Bowles and Gintis (2004)

<sup>3</sup> Uslaner (2002), for example, finds a small but positive correlation between generalized and particularized trust. See his discussion in chapter 2, particularly pp. 32–33.

<sup>4</sup> Several studies have assessed confidence across racial/ethnic lines using experiments developed in economics; but the results are mixed. Eckel and Wilson (2004) report pronounced racial differences in trust in the United States, whereas Glaeser et al. (2000) find little difference in trust, but a large gap in trustworthiness (i.e., the willingness to reciprocate). Research conducted in Israel finds that male European- and Eastern-Israelis both distrust Easterners, but women show no particular bias (Fershtman and Gneezy 2001).

are conceptually distinct, and we expect them to have independent effects on cross-ethnic confidence.

Generalized trust, as Uslaner (2002) argues, is the willingness to consider strangers as part of one's moral community; to assume that they share fundamental values at some level. It appears to reflect a person's basic optimism about life and the ability to influence it; satisfaction with life; and sense of equality (cf. Smith 1997; Uslaner 2002). It varies with changes in a person's life circumstances or with broader changes in society. But studies of generational differences also show that new cohorts begin their adult life with very different levels of faith in others, and that the differences persist over time (Putnam 2000; Robinson and Jackson 2001; Uslaner 2002). Trust thus seems to have a stable component from adolescence through the life cycle.

Broad-gauged trust should also include confidence developed in more concrete contexts, in everyday transactions such as those with neighbors and coworkers. Thus, people with low levels of generalized trust might still develop faith in others within a relatively familiar environment. We label this as "intermediate" trust, to distinguish it from faith in closer contacts such as family and friends, on the one hand, and generalized faith in people on the other.<sup>5</sup>

Broad-gauged trust should include confidence in government as well. Political institutions can reduce the risk of opportunism in transactions across ethnic lines, and help to reduce the uncertainty that can arise in dealing with perceived outsiders (Levi 1996; Posen 1993; Weingast 1998). People who see government as providing stable rules of the game should be more willing to trust outside their own ethnic group.

The second component of our model is the disposition to trust other ethnic groups, given one's own ethnic attachment. This would include adherence to what Hardin (1995) labels norms of exclusion, the informal rules that help to define who is "one of us" and who is an outsider. Values that reinforce one's sense of difference should diminish the willingness to trust outgroups. Having a sense of ethnically based bias should also influence cross-ethnic confidence. People who believe they have experienced discrimination, or who see their group as systematically disadvantaged, should display less confidence in other ethnic groups.

The third component in our model is outgroup specific. The willingness to trust a given outgroup should depend on a person's stereotypes about and contact with that group. Stereotypes of warmth or likability and competence appear to be especially relevant. As Fiske (2000, 2002) suggests, these perceived traits help to define who may be a friend and who is a rival.

Contact with outgroup members in everyday settings should also be a key element underlying cross-ethnic trust. Day-to-day contact can help to provide information about the outgroup, especially information to individualize its members. As Pettigrew (1998) notes, it

can also simply habituate people to being in a mixed environment and thus reduce anxiety about being face to face with a stranger. Of course, one might also argue that contact can breed competition and hostility; but mounting empirical evidence suggests that in nonconflict settings, the effect is generally positive rather than negative (see, e.g., Oliver and Wong 2003; Pettigrew and Tropp 2000; Welch et al. 2001).<sup>6</sup>

We are skeptical, then, of the idea that ingroup trust necessarily comes at the expense of faith in other ethnic groups or in others generally. And we are skeptical of the idea that low generalized trust automatically implies a lack of confidence among ethnic groups. We assess these issues first with evidence on different types of trust, then evaluate how well our model helps to explain cross-ethnic confidence. We also examine how ingroup trust and outgroup trust are connected.

If assumptions about exclusionary trust are valid, we should find a negative correlation between ingroup and generalized confidence in others; and a negative correlation between ingroup trust and outgroup trust. We should also find little to no difference in an individual's confidence toward different outgroups. If our argument is correct, generalized trust should be positively correlated with faith in one's own and in outgroups. And an individual's confidence in another group should depend on the outgroup in question.

## TRUST AND ETHNICITY IN RUSSIA

The Russian Federation offers a good context for evaluating our argument, with a low reported level of generalized faith in others, and widespread particularized trust. Surveys fielded during the 1990s reported that 25% to 30% of the population displayed generalized trust (Dowley and Silver 2002; Gibson 2001; Rose 1995). But the surveys revealed far higher faith in acquaintances, and even more trust in friends and family (Gibson 2001; Rose 1995).

The combination of low generalized and high particularized confidence in others implies fertile ground for exclusionary ethnic trust—especially where ethnic differences have been so highly politicized. The surge of ethnic assertiveness in the late 1980s through the end of the 1990s opened up extended public debate about the role and rights of ethnic minorities. And Russia's ethnic republics set out to promote ethnic revival for their titular nationalities, with expanded use of titular languages, renewed interest in groups' history and culture, and in some cases renewal of traditional religions as well.

We focus here on two republics, Tatarstan and Sakha-Yakutia, that experienced substantial ethnonational mobilization in the late 1980s and 1990s (Balzer and

<sup>5</sup> This overlaps in part with what Uslaner (2002) labels "strategic" trust, based on transactions with specific individuals. But we want to emphasize that the experience can be generalized beyond specific individuals within the local community, and thus be cumulative.

<sup>6</sup> Two other questions also arise in evaluating the contact hypothesis: self-selection and context. Where these effects have been addressed explicitly, the evidence suggests that contact can have a positive effect even controlling for self-selection (Oliver and Wong 2003; Powers and Ellison 1995), and that it can do so in a variety of different contexts (Pettigrew 1998; Pettigrew and Tropp, 2000). However, more favorable contexts (e.g., where groups are relatively equal) heighten the effect (Pettigrew and Tropp).

Vinokurova, 1996; Gorenburg 2003). These two were among the leaders in asserting their rights, and devoted considerable resources to reviving the titular languages and cultures.

Both regions are also ethnically diverse. Tatars make up around half of the population in Tatarstan; and Yakuts, roughly 40% in Sakha. Russians in both cases make up another 40% to 45%. The two major groups in each republic are large enough to come into regular contact and also into potential competition and conflict.<sup>7</sup>

The two regions also offer some important cultural and socioeconomic contrasts that may affect cross-ethnic trust. The titular languages and Russian belong to different language families (Tatar and Yakut are both Turkic).<sup>8</sup> With respect to religion, most Russians identify themselves as Orthodox, and most Tatars identify themselves as Muslims. Fewer Yakuts identify with a religion, and of those who do, most identify with either Orthodoxy or shamanism. A third distinction involves race—Yakuts are Asian, whereas Tatars and Russians are Caucasian. A final distinction lies in the relative socioeconomic status of the two dominant ethnic groups in each republic. Although Tatars and Russians are more or less equal in the distribution of occupations and incomes, there is a larger gap between the predominantly rural Yakuts and the predominantly urban Russians in Sakha-Yakutia (Bahry 2002).

Given these conditions, we should find high trust in ingroups and little to none in outgroups. Evidence of trust in both should thus be all the more compelling.

## DATA AND ANALYSIS

We concentrate here on the level of interethnic trust between the titular nationality and the Russians in each republic. Our data come from a survey conducted in Tatarstan and Sakha-Yakutia in spring–summer 2002, with republic-wide probability samples. Questionnaires were developed in Russian and translated into Tatar and Yakut. The Tatar and Yakut versions were then blind-backtranslated to ensure linguistic equivalence. Interviews with members of the titular nationality and with Russians in almost all cases were conducted by same-nationality interviewers. Titular-nationality respondents had the option of using either the titular language or the Russian language in the interview.

All told, 2,572 respondents were interviewed. Response rates were 81% in Tatarstan and 72% in Sakha. Ten percent of the interviews were verified *ex-post* by independent staffers (except in very small villages).

<sup>7</sup> But cross-ethnic contact is more common in Tatarstan, with its more compact territory, larger cities, and more urbanized population. In Sakha, most Yakuts live in villages, whereas the vast majority of Russians live in the cities. Sakha also has fewer and smaller cities, and more limited infrastructure linking them to villages.

<sup>8</sup> And the titular nationalities generally speak both, whereas very few Russians know the titular language. Among Russians in Tatarstan, 2% say they can speak Tatar freely or well, and around 14% say they can speak it poorly. The corresponding figures for Russians in Sakha are 2% and 7%, respectively.

Additional details on the survey design and sample are provided in Appendix 1. Details on the variables used in the analysis are found in Appendix 2.

Our primary dependent variable is the level of interethnic or *Outgroup trust*. For the two titular groups, Tatars and Yakuts, this variable comes from a question about the level of trust in Russians. For Russians, the question is about the level of trust in the titular group in each republic. We think that cross-ethnic confidence should be a function of:

*Generalized trust*, based on the standard question of whether most people can be trusted, or “you can’t be too careful” [“you always need to be careful”] in dealing with other people (cf. Alesina and La Ferrara 2002; Gibson 2001; Knack and Keefer 1997; Uslaner 2002);

*Political trust*, an average score from three questions on the level of confidence in federal, republic and local governments;<sup>9</sup>

*Intermediate trust*, from two items about the level of confidence in one’s neighbors and coworkers;

*Outgroup stereotypes*, a mean score for four word-pair items rating the titular group and Russians;

*Intergroup contact*, from two questions about the ethnic mix among an individual’s neighbors and coworkers;

*Ingroup norms*, an average of three questions asking how important it is for a “true” member of one’s ingroup to follow certain behaviors—sending children to one’s native-language school; marrying within the group; and speaking only the group’s native language.

*Individual discrimination*, based on a question about an individual’s personal experience with discrimination due to nationality; and

*Collective discrimination*, based on a question about the importance of nationality in access to good jobs in the republic. (We include these separately because they may operate in somewhat different ways.)

We also include several control variables. Residence in a village might diminish confidence across group boundaries, in part because villages tend to be more homogeneous and more insulated from the larger society. Residence in the largest cities might have the opposite effect. Age cohort might exert an impact, because younger people especially have been the focal point of republic government efforts to reinvigorate the titular language and culture. And education might prove significant as well, by increasing tolerance and thus trust.

Given the different histories and different statuses of the titular groups and Russians in each republic, we think it is important to conduct the analysis separately for each of them. To determine group membership,

<sup>9</sup> Initially, we ran our analysis with trust in the federal, republic, and local governments separately, because it seemed likely that people in sovereignty-manded regions might have divergent views of the federal versus republic and local governments. But the results showed that trust in all three levels was highly correlated, with an interitem alpha of .79, so we combined them. We also tried using measures for trust in leaders (Putin and republic Presidents Shaimiev and Shtyrov) rather than institutions. The results were very similar to those reported here.

**TABLE 1. Generalized and Particularized Trust by Ethnic Group**

	Tatarstan		Sakha-Yakutia	
	Tatars	Russians	Yakuts	Russians
Percentage with Generalized Trust in People				
Most people can be trusted	17.9	19.9	17.6	22.6
"Depends on person/situation"	20.0	22.2	14.3	16.0
Always need to be careful	62.1	57.9	68.1	61.3
Intermediate Trust				
Trust neighbors completely/somewhat	83.7	82.3	82.8	76.1
Of which:				
Completely	31.6	28.4	37.4	28.3
Somewhat	52.0	53.9	45.4	47.9
Trust coworkers completely/somewhat	76.9	77.2	88.7	80.6
Of which:				
Completely	18.2	16.0	34.5	24.1
Somewhat	58.7	61.3	54.2	56.2
Ingroup Trust	95.1	97.0	90.7	86.7
Of which:				
Completely	40.6	24.2	32.5	22.9
Somewhat	54.5	72.8	58.2	63.8
Outgroup Trust	91.7	85.7	69.5	71.2
Of which:				
Completely	25.8	16.0	10.2	15.3
Somewhat	65.9	69.7	59.3	55.9
Mean Difference in Trust Between In- and Outgroup	0.19	0.23	0.49	0.31
Number of cases (range)	528–614	426–544	524–551	535–574

Note. For questions and definitions of each variable, see Appendix 2. "Mean difference in trust" is calculated as ingroup minus outgroup trust (high = trust ingroup more).

we rely on an open-ended question about a person's subjective ethnic identity.<sup>10</sup>

## THE DIMENSIONS OF INTERETHNIC TRUST

As in earlier research on Russia, we find that most people are cautious about others. A little more than 60%, on average, say that you "always need to be careful" in dealing with others (see Table 1). Around 20% say that most people can be trusted, and a roughly similar proportion responds that "it depends." Generalized trust, then, is low.

Also as in earlier research, we find higher levels of particularized trust. Between 86% and 97% of respondents trust their ingroup to some degree.<sup>11</sup> Interme-

diate trust is high as well. Over three fourths of our sample trust their neighbors; and a similarly high percentage trust their coworkers.

If we stopped here, we would have told the usual story of zero-sum trust: there is little generalized faith in others, and instead people turn inward to trust their ethnic group. What is striking, however, is that people display considerable confidence in the outgroup. Thus, over 90% of Tatars trust Russians; and local Russians express almost the same high level of faith in Tatars. Mutual confidence is somewhat lower among Yakuts and Russians, but still registers at around 70% in each case (see Table 1).

What is also striking is that the correlations among generalized, in- and outgroup trust are all positive (Table 2). Thus, faith in one's own and faith in the major outgroup are not mutually exclusive, but complementary.

It could be, of course, that the relatively high level of mutual confidence simply reflects social desirability (cf. Javeline 1999). People may be unwilling to express negative sentiments about an outgroup, especially when face-to-face with an interviewer of another nationality. But virtually all titular and Russian respondents were interviewed by same-nationality interviewers. And other evidence in the survey shows that people

<sup>10</sup> And most respondents named a single nationality. The survey also asked if people identified with any other nationality as well; but relatively few people did so. Among people who identified themselves primarily as Tatars, only 3.2% also identified as Russians. Among primary Russian identifiers in Tatarstan, only 1.4% also identified as Tatars. Among Yakuts and Russians in Sakha, the corresponding percentages are 3.2% 4.0%, respectively. (A small percentage in each case also identified with other groups—for example, Russians with Ukrainians, or Tatars with Chuvash or Bashkirs.) Nor were there very many respondents from titular Russian families of origin: fewer than 10% of Tatars, Yakuts or Russians were children of such families.

<sup>11</sup> We combine the responses of "trust completely" and "trust somewhat" in order to allow comparisons with earlier research (cf. Gibson, 2001). We have also found from other questionnaire items that some respondents (often the older and less educated) favor more

categorical answers in general, whereas others favor less categorical ones. However, our multivariate analysis below (Table 3) treats "trust completely" and "trust somewhat" as separate categories.

**TABLE 2. Correlations between Generalized and Particularized Ethnic Trust**

	Tatarstan		Sakha-Yakutia	
	Tatars	Russians	Yakuts	Russians
Generalized * ingroup trust	.07*	.16***	.09**	.13***
Generalized * outgroup trust	.17***	.20***	.17***	.19***
Ingroup * outgroup trust	.66***	.53***	.34***	.44***
Number of cases (range)	579–588	515–521	537–543	534–555

*Note* The data are simple correlation coefficients (Pearson's *R*). \*\*\* Significant at  $p < .01$ . \*\* Significant at  $p < .05$ . \* Significant at  $p < .10$ .

did not simply give “rote” positive responses about interethnic relations. One indication is that responses on outgroup trust vary substantially across different outgroups. People answered more readily when asked about the major, visible outgroup (titular or Russian) than when asked about less visible or more distant groups. The “don’t know’s” for questions about trust in titular groups and in Russians ranged from 1% to 7%. Questions about trust in other, less visible or proximate groups (Chechens, Chinese, Jews, and Americans) elicited “don’t know’s” from up to 50% of respondents. And among people who did give a substantive answer on these latter groups, levels of trust varied much more.<sup>12</sup> (Given the high rates of item nonresponse, we evaluate trust in the less visible/proximate groups separately, later.)<sup>13</sup>

People also varied in their responses when asked to rate the major outgroup (titular or Russian) on several stereotypical traits. Although most people held relatively neutral or positive images of the other group, their images varied depending on the trait in question. Few people simply gave rote responses.

In addition, respondents seemed little inclined to endorse exclusionary ingroup norms just to please interviewers. Figure 1 plots the mean level of importance that people attached to three different norms: marrying within one’s own group, speaking only the ingroup language, and having one’s child educated in that tongue. Most people in each of our four groups saw these behaviors as less than essential to be a “true” member of their group.<sup>14</sup>

<sup>12</sup> Of the people who gave a substantive answer about trust in less visible/proximate groups, 27% to 49% expressed trust in Jews; 16% to 34% in Chinese; 20% to 38% in Americans, and 11% to 24% in Chechens. Factor analysis of all six ethnic trust questions yields two factors—one for trust in the titular nationality and Russians and a second factor for trust in these latter nationalities. We think this reflects a difference between proximate and more distant groups.

<sup>13</sup> Unless otherwise noted, we exclude “don’t know” responses from the analysis.

<sup>14</sup> One reason might be that ethnicity is simply not very salient to individuals. But other items suggest that it is salient, especially for titular groups. On questions tapping the importance of various sources of group identification, nationality ranked second only to family, and higher than eight other sources of affiliation (generation; standard of living; resident of the republic, of Russia; or of the locality, religion, place of origin, and occupation), for titular groups and Russians alike. When asked if they seldom thought about their nationality or always thought about it, over 60% of Tatars and over 75% of Yakuts said “always,” whereas 30% to 40% of Russians said the same.

## IDENTIFYING THE ROOTS OF INTERETHNIC TRUST

Our model holds that outgroup trust should depend on orientations at three levels. One is broad-gauged trust, which would dispose people to trust others regardless of their group affiliation. The components of broad-gauged trust include generalized faith in others; intermediate trust (in neighbors and coworkers), and confidence in government. The second level would dispose people to trust other ethnic groups on the whole. In this case, trust would depend on attachment to norms of ethnic exclusion and on an individual’s sense of being the victim of ethnic discrimination. The third level would dispose people to trust in specific outgroups, based on stereotypes of the outgroup and on contact with its members.

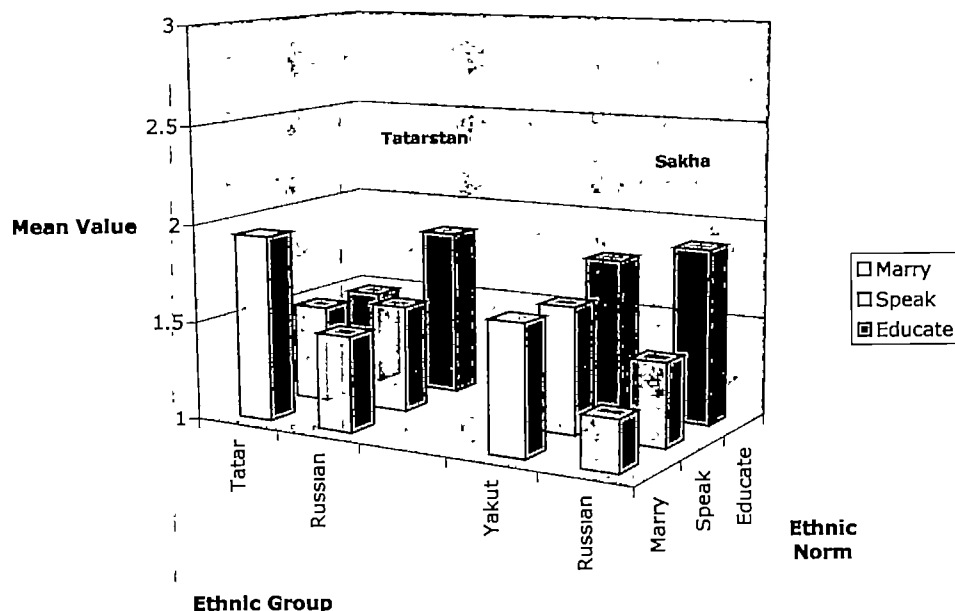
We evaluate these arguments here with an ordered logit (Table 3). The results bear out much of our argument.<sup>15</sup> All three components of what we have termed broad-gauged trust prove significant, almost across the board. Intermediate trust—in neighbors and coworkers—leads to higher faith across ethnic lines. So, too, does confidence in government. The more positive the view of the authorities, the more trust in the major outgroup. And generalized faith in people increases cross-ethnic trust for three of our four groups.

Also as we would predict, support for exclusionary ethnic norms lowers confidence across ethnic lines. People who endorse endogamy, exclusive use of their own language, and exclusive own-language education for their children have less faith in the outgroup.

Attitudes toward the particular outgroup also prove significant. The more the outgroup is viewed as likable and competent, the greater the confidence across group boundaries. Contact, however, appears to play a more limited role. It proves significant in one republic (Sakha) but not in the other, and moreso for the titular group than for Russians. We would need a more detailed study to determine why these results differ; but one possible explanation is that interethnic contact is less common in Sakha than in Tatarstan, and the effect may thus be heightened.

<sup>15</sup> The control variables—urban/rural residence, age and education, generally have little impact. When these variables alone are used to predict levels of interethnic trust, the pseudo- $R^2$ s are .01, .02, .03, and .05, for Tatars and Russians in Tatarstan, Yakuts, and Russians in Sakha, respectively.



**FIGURE 1. Mean Support for Ethnic Norms by Major Ethnic Groups within Each Republic**

Note. Marry: "How important is it to marry a same-nationality spouse?"  
 Speak: "How important is it to speak only the language of one's own nationality?"  
 Educate: "How important is it to send a child to an own-language school?"  
 For each item, the scale is 1 = not important; 2 = desirable; 3 = essential.

**TABLE 3. Sources of Interethnic Trust**

	Tatarstan		Sakha-Yakutia	
	Tatars (Trust in Russians)	Russians (Trust in Tatars)	Yakuts (Trust in Russians)	Russians (Trust in Yakuts)
Size of settlement				
Big city	.16	.35	-.28	.07
Village	-.44	-.09	-.55	-.24
Age				
Born 1940 or earlier	.14	.32	-.59*	-.11
Born 1970 or later	-.22	.06	-.04	-.81***
Education	.25	-.14	.12	-.16
Sense of discrimination				
Individual	-.03	-.94**	-.48	.18
Collective	-.06	-.10	.06	-.18
Generalized trust	.24*	.54***	.30**	.05
Trust neighbors/coworkers	.40**	.71***	.63***	.60***
Trust government	.74***	.62***	.87***	.62***
Positive stereotypes of other group	.36***	.66***	.40**	.82***
Exclusionary ethnic norms	-.31*	-.41*	-.89***	-.65***
Contact with outgroup	.00	.19	.27**	.22*
Intercept 1	-.89	-1.27	-.15	-1.14
Intercept 2	.85	.83	1.84*	.88
Intercept 3	4.71***	5.26***	5.56***	4.31***
Pseudo R <sup>2</sup>	.16	.32	.29	.35
-2 LL	721.54	545.67	907.87	884.09
Number of cases	453	365	488	463

Note: The dependent variable is based on the question "How much do you trust [name of other nationality]—completely, somewhat, distrust somewhat, or distrust completely?" The question here is treated as a 4-point scale, where "trust completely" is high. The numbers are ordinal logit coefficients. For definitions of variables, see Appendix 2. \*\*\* Significant at  $p \leq .01$ . \*\* Significant at  $p \leq .05$ . \* Significant at  $p \leq .10$ .

**TABLE 4. Inclusionary and Exclusionary Ethnic Trust**

	Tatarstan		Sakha-Yakutia	
	Tatars	Russians	Yakuts	Russians
Percentage who are:				
<i>Inclusionary</i> (trust both in- and outgroup)	90.8	85.7	67.0	68.4
<i>Exclusionary</i> (trust ingroup only)	4.4	11.2	23.3	18.9
<i>Alienated</i> (trust outgroup only)	0.9	—	2.6	2.8
<i>Atomized</i> (trust neither)	3.9	3.1	7.1	9.9
Number of cases	588	518	537	535

Perceptions of discrimination have even less impact, proving significant only for one of the four groups, and then only for individual-level perceptions of bias (among Russians in Tatarstan). One reason might be the inclusion of two potentially related measures of discrimination in the same model. But including each one in the model separately yields the same results as those listed in Table 3.<sup>16</sup>

All told, then, the results generally support our model of cross-ethnic trust. They show, moreover, that there are some strong similarities in the roots of trust across the four groups in our study, despite their religious, linguistic, and socioeconomic differences.<sup>17</sup> But the results are not identical, as our findings on contact and perceived discrimination attest.

### Gauging Inclusionary versus Exclusionary Trust

Our analysis thus far also raises some doubts about the prevalence of exclusionary ethnic trust. High levels of intra- and interethnic confidence suggest that relatively few people are “zero-sum” trusters. We can determine their numbers more directly, however, by combining questions on in- and outgroup trust. These yield a four-fold typology:

- “inclusionary” (people who express confidence in both the in- and outgroup);
- “exclusionary” (people who trust their own, but distrust the other group);
- “alienated” (people who trust the other group but distrust their own); and
- “atomized” trust (people who distrust both the in- and the outgroup).

<sup>16</sup> Because other independent variables are also related, we ran a variety of tests to determine how the correlations might influence our conclusions about the factors shaping outgroup trust. We partialled the impact of generalized trust out of our measures of confidence in government and intermediate trust; but the signs and levels of significance of the latter two variables remained the same as those in Table 3. We also partialled outgroup stereotypes out of our measure of ingroup norms, and the results remained the same. Finally, we estimated a two-stage least-squares model to check our assumptions about the impact of generalized trust in Table 3. (It could be argued that outgroup trust shapes generalized faith in others, rather than the reverse.) But the results confirmed the analysis in Table 3.

<sup>17</sup> We also ran the analysis in Table 3 with additional variables to tap commitment to a group's traditional religion and use of the group's language, but neither proved to be significant.

The frequencies are presented in Table 4. Exclusionary trusters constitute a minority in each case, from 4% among Tatars to 23% among Yakuts. Inclusionary trust is more common. Two thirds of respondents in Sakha and four fifths of respondents in Tatarstan express some confidence in both their own and the other group.<sup>18</sup> Another small minority (from 3% to 10%) are atomized, trusting neither group. And only a handful are alienated.

We explore what differentiates people in each category with an unordered logit (Table 5).<sup>19</sup> Inclusionary trust—in both one's own group and in the major outgroup—is the omitted category. (The “alienated” were dropped from the analysis because there were too few cases.) For each ethnic group, the data in column 1 help identify the factors that lead to more exclusionary (rather than inclusionary) trust; and in column 2, the factors that lead to atomization (rather than inclusion).

With respect to exclusionary trust, one factor stands out as significant across all four groups: stereotypes. The more negative the perception of outgroup members, the more the zero-sum trust. Ingroup norms have a slightly less consistent impact, as does intermediate faith in others. And as in our earlier analysis (Table 3), contact proves significant for the two major groups in Sakha. Perceived discrimination also has a selective effect for Russians in each region.

Atomization—distrust of both in- and outgroup—appears to be associated chiefly with distrust of government. Intermediate trust also plays a role, though the impact is less uniform. Otherwise, the roots of atomization vary more. But given the small number of cases in this category, we would not want to read too much into these differences.

Thus, we do find some evidence of exclusionary ethnic trust, and it is connected both to ingroup bias and to

<sup>18</sup> We also calculated a differential trust score, by subtracting outgroup from ingroup trust. Seventy percent of respondents received a score of zero—that is, they expressed the same degree of trust in their own and in the outgroup. Roughly 26% expressed more trust in their own group, and another 3% expressed less. An Ordinary Least Squares regression analysis with the same independent variables as those in Table 3 shows that exclusionary ingroup norms and stereotypes of the outgroup are the most significant factors predicting differential trust.

<sup>19</sup> Because of the small numbers in some categories, and because most of the background variables in the analysis have relatively little effect compared to the subjective ones, we dropped age, place of residence, and education from this part of the analysis.

**TABLE 5. Sources of Exclusionary and Inclusionary Trust by Major Ethnic Groups within Republics**

	Tatars		Russians in Tatarstan		Yakuts		Russians in Sakha	
	Exclusionary (Trust Ingroup Only)	Atomized (Trust Neither)	Exclusionary (Trust Ingroup Only)	Atomized (Trust Neither)	Exclusionary (Trust Ingroup Only)	Atomized (Trust Neither)	Exclusionary (Trust Ingroup Only)	Atomized (Trust Neither)
Generalized trust	-.15	-1.63*	-.20	-.49	-.26	-.22	-.07	-.10
Trust neighbors/ coworkers	-1.01***	-.71*	-.37	-.60	-.60**	-.87***	-.37*	-.70***
Trust government	.05	-1.89***	-.32	-1.77***	-.31	-1.27***	-.50**	-.95***
Exclusionary ethnic norms	.37	.49	.72**	-1.08*	1.26***	-.03	.84***	.20
Positive stereotypes of outgroup	-.79***	-.06	-.68***	-.74**	-.31**	-.20	-.99***	-.72***
Contact with outgroup	-.38	-.67**	-.15	-.67	-.41***	-.19	-.56***	-.16
Perceived individual discrimination	-.83	-.76	1.17***	1.87***	.36	.21	-.08	.29
Perceived collective discrimination	.42	.25	.44	1.34	.13	.11	.27*	.22
Intercept	.67	6.07	-3.46**	-.67	-.16	3.84***	.07	1.60
Pseudo R <sup>2</sup>	.27		.35		.27		.32	
-2 log likelihood	247.7		302.5		628.4		600.7	
Number of cases	464		391		477		453	

Note. Unordered multinomial logit, where the dependent variable has three categories: 1, trust neither in- nor outgroup; 2, trust ingroup only; and 3, trust both. The base category here is "trust both." (Those who are alienated—people who trust the outgroup only—were omitted due to the very small number of cases.)

negative views of the outgroup. What is most striking, though, is that it is far from the modal category. More people are inclusionary.

### CROSS-ETHNIC CONFIDENCE IN MORE DISTANT GROUPS

To this point, we have examined the roots of cross-ethnic trust between relatively large and visible groups, who are most likely to be in contact on a daily basis. How would our model fare when the question turns to trust in other, less proximate or visible groups? As we noted earlier, both salience and trust decline with social and physical distance. Questions about trust in Jews, Chinese, Americans, and Chechens evoke many more "don't know" responses than do questions about titular groups and Russians. And when people do give a substantive reply, they express lower levels of confidence in less proximate groups.

But the underlying roots of cross-ethnic confidence in less proximate groups appear to be very similar to those identified in Table 3. We cannot replicate the analysis in Table 3 in its entirety, because our data do not include outgroup-specific questions on stereotypes of, or contact with, each of the less proximate groups. We can, however, estimate the effects of broad-gauged trust, exclusionary ethnic norms, and perceived discrimination. We thus ran an ordered logit similar to that in Table 3 to estimate how much titular groups and Russians in each republic trust each of the less proximate groups.<sup>20</sup> Because this produced 16 equations—four

analyses for each titular group and for Russians in each republic—we summarize the results here.

Three factors turn out to be significant almost across the board (in 12 or more of the 16 equations). People who have higher confidence in government, more generalized faith in people, and less attachment to ingroup norms express more trust in less proximate outgroups. Intermediate trust, on the other hand, seems to have much less impact here than it does for trust between titular groups and Russians. Trust in neighbors and coworkers thus appears to be capturing the effect of living in what is mostly a biethnic context.

### CONCLUSIONS

The value of interpersonal trust is now a central element in theories of democracy and markets. But the benefits seem to stop where ethnic attachments begin. A growing literature thus characterizes ethnic capital as the inverse of more general social capital, and of outgroup trust.

Our analysis suggests that this image may be too limited. High ingroup trust is no barrier to faith in others. In fact, we found most people to be inclusionary—displaying confidence both in their own and in the major outgroup. We are not suggesting that the same proportions would hold elsewhere. In fact, our argument is that we should not prejudice the proportions, because they will depend on levels of what we have termed broad-gauged trust, attachment to one's

<sup>20</sup> We also did an analysis of the "don't know" responses for questions about trust in each of the four less proximate groups. "Don't

knows" were more prevalent among older, less educated, and rural respondents. They were not very closely related to our other measures of trust, to ingroup norms, or to a sense of ethnic discrimination.

ingroup, and stereotypes of, and contact with, the particular outgroup.

We find strong support for this model of the roots of cross-ethnic trust. Broad-gauged trust leads to greater confidence across ethnic lines. In contrast, strong attachment to ingroup norms and negative stereotypes of the outgroup lower cross-ethnic confidence. These findings are similar, moreover, among ethnic groups—Tatars, Yakuts, Russians—with marked differences in language, religion, and culture. Everyday contact appears to have a more selective effect, stronger where contact itself is more limited.

Many of the sources of interethnic trust also appear to be similar when people are asked about their level of trust in other, less proximate groups as well. Generalized faith in others figures almost across the board, as do confidence in government and attachment to ingroup norms. Our analysis shows, however, that people display more trust in, and are more likely to give an opinion about, groups they see on a day-to-day basis.

Thus, trust varies across different outgroups, depending at least in part on contact and familiarity. But our data suggest an important distinction between lack of contact/familiarity (and high rates of “don’t know” responses) versus distrust. They also suggest the possibility that the effect of contact on individuals’ willingness to trust outgroups may be u-shaped: limited both where transactions across ethnic lines are very common and where they are very scarce.

Our results also hold several other implications for the study of trust. One is about the distinction between generalized and particularized confidence in others. As in other studies of Russia, our analysis finds low generalized and high particularized confidence in others. But we also find intermediate forms of trust that do not fit neatly into either category. Relatively high levels of confidence in neighbors and coworkers, and in major outgroups, suggest that faith in others is somewhat broader than many images of the particularized variant alone would suggest. These forms of trust seem to depend in part on familiarity; but they extend to whole groups and not just to the individuals people see face-to-face.

A second implication relates to the role of confidence in government. We find, as several other authors do, that faith in political institutions bolsters cross-ethnic trust. But the Russian government has been rated as increasingly undemocratic since the late 1990s. This all implies that the key feature connecting confidence in government to cross-ethnic trust need not be the degree of democracy or transparency, as some authors suggest. It may simply be the provision of stable rules of the game (cf. Barber 1983; Posen, 1993; Sztompka, 1999).

Finally, our results should make us skeptical of generalized faith in others as an indicator of confidence across ethnic lines. As we have demonstrated, it does *contribute* to cross-ethnic trust. But the two are clearly not interchangeable. Only one fifth of our sample displays generalized faith in others; but roughly four fifths trust the major outgroup in their republic. Trust is also higher for some less proximate ethnic groups than it

is for “people” in general. If so, then arguments about the connection between ethnicity and trust need to be refined.

## APPENDIX 1: DESCRIPTION OF THE SURVEY

Our data come from a survey conducted in Tatarstan and Sakha in the summer and fall of 2002. The 2-hour, face-to-face survey covered a number of issues ranging from work to social relations and ethnic identification to trust. The questionnaire was developed in English and Russian with collaborators from Demoscope in Moscow, and then translated into Tatar and Yakut. The Tatar and Yakut versions were subsequently blind-backtranslated to ensure linguistic equivalence. In almost all cases, titular and Russian respondents were interviewed by same-nationality interviewers. Titular-nationality respondents were interviewed by bilingual interviewers and could opt to give the interview either the titular language or Russian. The eligible population included non-institutionalized permanent residents 18 years of age and older.

The stratified, random sample was designed to achieve two goals—to allow comparisons between the titular nationality and Russians in each republic and to allow inferences about the populations of each republic as a whole. However, comparisons across ethnic groups could be complicated by the fact that the two groups were unevenly distributed, with Tatars and Russians making up around 48% and 43% of Tatarstan’s population, respectively, and with Yakuts and Russians accounting for approximately 40% and 45% in Sakha. The survey thus included an oversample of the underrepresented nationality in each case.

The sample design began with two strata: urban and rural, using data from the most recent census updates. Urban areas were then further stratified by size and drawn randomly with probability proportional to size. Each urban area in the sample was then partitioned into districts, and districts were randomly selected for inclusion in the sample. Within each sample district, a list of all dwellings was constructed by visual inspection and consultation with authorities. In the case of dormitories and communal apartments, each room or space housing a separate household was treated as a dwelling unit, not the entire building. Then, a number of dwelling units were selected systematically starting with a random number. An individual from each drawn household was then selected using the Kish (1965) procedure.

Sampling in rural areas presented more problems. Cost considerations and lack of detailed data made it impossible to build a sample from the ground up based on all rural settlements, or to conduct only a few interviews apiece in widely dispersed settlements. As an alternative, villages were drawn from the raions (regions) included in the urban stratum. However, many villages in the two republics are ethnically homogeneous, and because only a few villages could be selected from most raions, we wanted to avoid the problem of disproportionate representation of one or the other major ethnic group. As a result, within each selected raion, villages were stratified by ethnic composition—predominantly (90% or more) titular, predominantly (90% or more) Russian, and mixed. Within each of these strata, villages with at least 100 residents were ordered by size and selected by probability proportional to size. Within each village, households were selected systematically starting from a random number and drawn from the official residence registration book (*pokhoziastvennaia kniga*). The Kish procedure was employed to select an individual respondent from each drawn household. As

in urban areas, at no point did interviewers exercise discretion in the selection of households or respondents.

Geography and budgetary considerations imposed some limitations on the sample design. Remote areas of Sakha were eliminated because many points were sparsely settled, and not regularly accessible by scheduled transportation. Such exclusions are common in national surveys employing face-to-face interviews; certain territories are eliminated in advance due to practical considerations such as very low population density, low accessibility, or political unrest. (Thus, in the United States, Alaska and Hawaii are typically excluded from national samples.)

A total of 2,572 people were interviewed: 1,266 in Tatarstan and 1,306 in Sakha. Response rates were 81% and 72%, respectively. Ten percent of completed questionnaires were chosen for inspection of interviewers' work by independent evaluators from Moscow, although these inspections were not typically conducted in very small villages. Here, we exclude 59 respondents who were participants in another part of the study and who did not fall into our original sample.

## APPENDIX 2. VARIABLES USED IN THE ANALYSIS

*Respondent nationality:* "What is your nationality?" [responses were open-ended].

*Generalized trust:* "Do you think that most people can be trusted, or that you always need to be careful in dealing with others?" [1 = always need to be careful; 2 = "depends" (volunteered response); 3 = most people can be trusted.]

*Ingroup trust:* "How much do you trust [Tatars/Yakuts/Russians]—completely, somewhat, distrust somewhat, or distrust completely?" [1 = distrust completely; 2 = distrust somewhat; 3 = trust somewhat; 4 = trust completely]. For Tatars, this question refers to Tatars; for Yakuts to Yakuts, and for Russians, it refers to Russians.

*Outgroup trust:* "How much do you trust [Tatars/Yakuts/Russians]—completely, somewhat, distrust somewhat, or distrust completely?" [1 = distrust completely; 2 = distrust somewhat; 3 = trust somewhat; 4 = trust completely]. For Tatars and Yakuts, this question refers to trust in Russians; for Russians, it refers to trust in either Tatars or Yakuts.

*Outgroup stereotypes:* An average score of four questions about stereotypes of Tatars (asked only in Tatarstan), Yakuts (asked only in Sakha), and Russians (asked in both regions). For each group, the questions were of the following form:

"Let's talk about character traits that are typical of people of different nationalities. Here is a scale where one means that (Tatars/Yakuts/Russians) are mostly characterized by slyness, and seven means that they are mostly characterized by simplicity. Where would you place (Tatars/Yakuts/Russians) on this scale?" [1 = slyness; 7 = simplicity].

The other three scales were [1 = hardworking; 7 = lazy]; [1 = sharp-witted; 7 = slow-witted]; and [1 = respectful of other nationalities; 7 = disrespectful of other nationalities].

These items were rescaled from -3 to +3, with the negative trait at -3 and the positive trait at +3; and averaged together. (Respondents were included if they answered on at least three of the four items.)

*Trust government:* a composite of three questions—"How much do you trust the federal government?" "The government of the republic?" "The administration of this city/village?" [1 = distrust completely; 2 = distrust somewhat; 3 = trust somewhat; 4 = trust completely]. These three were averaged together to give an index of trust in government. (Respondents were included if they answered at least two of the three items.)

*Trust neighbors/coworkers:* an index of two items—"How much do you trust your neighbors?" "Your coworkers?" [1 = distrust completely; 2 = distrust somewhat; 3 = trust somewhat; 4 = trust completely]. These two were averaged together to give an index of trust in neighbors/coworkers. (People were scored on this index if they answered at least one of the two questions.)

*Exclusionary ethnic norms:* Derived from questions asked about Tatars (in Tatarstan only), Yakuts (in Sakha only), and Russians (in both regions). The questions were: "What, in your opinion, is obligatory, what is desirable, and what is not important in order to consider someone a true (Tatar/Yakut/Russian)?"

- (a) Marry a (Tatar/Yakut/Russian) [1 = not important; 2 = desirable; 3 = obligatory]
- (b) Speak only (Tatar/Yakut/Russian) [1 = not important; 2 = desirable; 3 = obligatory]
- (c) Send children to (Tatar/Yakut/Russian)—language school [1 = not important; 2 = desirable; 3 = obligatory]

*Size of settlement:*

Big city = 1 if city is greater than 100,000 residents; 0 otherwise

Village = 1 if population of 3,000 or fewer people; 0 otherwise.

*Education:* 1 = completed less than secondary; 2 = completed secondary; 3 = completed higher.

*Sense of individual discrimination:* "Have you personally had to experience a violation of your rights or opportunities due to your nationality?" 1 = no; 2 = yes.

*Sense of collective discrimination:* "Do you think that a person's nationality in this republic affects his chances to get the best jobs?" 1 = not; 2 = "it depends" (volunteered); 3 = yes.

*Contact with outgroup:* Derived from two questions:

"What is the national composition of the collective where you work?" [1 = only Tatar/Yakut; 2 = mostly Tatar/Yakut; 3 = about half Tatar/Yakut; 4 = mostly non-Tatar/Yakut; 5 = no Tatars/Yakuts].

"What is the nationality of your neighbors?" [1 = only Tatar/Yakut; 2 = mostly Tatar/Yakut; 3 = about half Tatar/Yakut; 4 = mostly non-(Tatar/Yakut); 5, no Tatars/Yakuts].

These were rescaled for Tatars, Yakuts, and Russians so that a high value indicates that most or all neighbors or coworkers are of the other nationality. We then averaged the two to obtain a measure of contact with the outgroup. (People were scored on this index if they answered at least one of the questions.)

## REFERENCES

- Alberto Alesina, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions" *The Quarterly Journal of Economics* 114 (November): 1243-84.
- Alberto Alesina, and Eliana La Ferrara. 2000. "Participation in Heterogeneous Communities." *The Quarterly Journal of Economics* 115 (August): 847-904.
- Alberto Alesina, and Eliana La Ferrara. 2002. "Who Trusts Others?" *Journal of Public Economics* 85 (August): 207-34.
- Donna Bahry. 2002. "Ethnicity and Inequality in Postcommunist Transition. Evidence from Russia's Republics" *Europe-Asia Studies* 54 (July): 673-99.
- Marjorie Mandelstam Balzer, and Iuliana Alekseevna Vinokurova. 1996. "Nationalism, Interethnic Relations and Federalism: The Case of the Sakha Republic." *Europe-Asia Studies* 48 (January): 101-20.
- Bernard Barber. 1983. *The Logic and Limits of Trust*. New Brunswick, NJ: Rutgers University Press.

- Samuel Bowles, and Herbert Gintis. 2004. "Persistent Parochialism. Trust and Exclusion in Ethnic Networks" *Journal of Economic Behavior and Organization* 55: 1–23.
- Paul Collier. 2001. "Implications of Ethnic Diversity." *Economic Policy* 32 (April): 129–66.
- Dora Costa, and Matthew Kahn. 2003. "Civic Engagement and Community Heterogeneity: An Economist's Perspective." *Perspectives on Politics* 1 (March): 103–11.
- Kathleen Dowley, and Brian K. Silver. 2002. "Social Capital, Ethnicity and Support for Democracy in the Post-Communist States" *Europe-Asia Studies* 54 (June): 505–28.
- William Easterly. 2001. "Can Institutions Resolve Ethnic Conflict?" *Economic Development and Cultural Change* 49 (July): 687–706.
- William Easterly, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions" *The Quarterly Journal of Economics* 112 (November): 1203–50.
- Catherine C. Eckel, and Rick K. Wilson. 2004. "Initiating Trust: The Conditional Effects of Sex and Race among Strangers." Working Paper, Political Science, Rice University, Houston, TX.
- James Fearon. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8: 195–222.
- James Fearon, and David Laitin. 1997. "Explaining Interethnic Cooperation," *American Political Science Review* 90 (December): 715–35.
- Chaim Fershtman, and Uri Gneezy. 2001. "Discrimination in a Segmented Society. An Experimental Approach." *The Quarterly Journal of Economics* 116 (February): 351–77.
- Susan T. Fiske. 2000. "Stereotyping, Prejudice and Discrimination at the Seam Between the Centuries: Evolution, Culture, Mind and Brain." *European Journal of Social Psychology* 30 (May/June): 299–322.
- Susan T. Fiske. 2002. "What We Know Now About Bias and Intergroup Conflict, the Problem of the Century." *Current Directions in Psychological Science* 11 (August): 123–28.
- Francis Fukuyama. 1995. *Trust. The Social Virtues and the Creation of Prosperity*. New York: The Free Press.
- Francis Fukuyama. 1999. "Social Capital and Civil Society." IMF Conference on Second Generation Reforms, Washington DC (November). Available at <http://www.imf.org/external/pubs/ft/seminar/1999/reforms/fukuyama.htm#1> (January 16, 2005).
- Jose Garcia-Montalvo, and Marta Reynal-Querol. 2002. "Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth." Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=394926](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=394926) (January 25, 2005).
- James Gibson. 2001. "Social Networks, Civil Society and the Prospects for Consolidating Russia's Democratic Transition" *American Journal of Political Science* 45 (January): 51–69.
- James Gibson. 2004. "Does Truth Lead to Reconciliation? Testing the Causal Assumptions of the South African Truth and Reconciliation Process" *American Journal of Political Science* 48 (April): 201–17.
- Edward Glaeser, David Laibson, Jose Scheinkman, and Christine Soutter. 2000. "Measuring Trust." *The Quarterly Journal of Economics* 115 (August): 811–46.
- Dmitry Gorenburg. 2003. *Minority Ethnic Mobilization in the Russian Federation*. Cambridge: Cambridge University Press.
- Mark Granovetter. 1973. "The Strength of Weak Ties." *American Sociological Review* 78 (May): 1360–80.
- Russell Hardin. 1995. *One for All. The Logic of Group Conflict*. Princeton: Princeton University Press.
- Mary Herring, Thomas Jankowski, and Ronald Brown. 1999. "Pro-Black Doesn't Mean Anti-White. The Structure of African-American Group Identity." *The Journal of Politics* 61 (May): 363–86.
- Debra Javeline. 1999. "Response Effects in Polite Cultures." *Public Opinion Quarterly* 63 (spring): 1–28.
- Leslie Kish. 1965. *Survey Sampling*. New York: John Wiley.
- Stephen Knack, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics* 112 (November): 1251–88.
- Janet Tai Landa. 1995. *Trust, Ethnicity, and Identity. Beyond the New Institutional Economics of Ethnic Trading Networks, Contract Law, and Gift-Exchange*. Ann Arbor: University of Michigan Press.
- Rafael LaPorta, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny. 1997. "Trust in Large Organizations." *American Economic Review Papers and Proceedings* 87 (May): 333–38.
- Margaret Levi. 1996. "Social and Unsocial Capital: A Review Essay of Robert Putnam's 'Making Democracy Work'." *Politics and Society* 24 (March): 48–55.
- J. Eric Oliver, and Janelle Wong. 2003. "Intergroup Prejudice in Multiethnic Settings." *American Journal of Political Science* 47 (October): 567–82.
- Thomas Pettigrew. 1998. "Intergroup Contact Theory." *Annual Review of Psychology* 49: 65–85.
- Thomas Pettigrew, and Linda Tropp. 2000. "Does Intergroup Contact Reduce Prejudice? Recent Meta-Analytic Findings." In *Reducing Prejudice and Discrimination*, ed., Stuart Oskamp. Mahwah, NJ: Lawrence Erlbaum Associates, 93–114.
- Barry Posen. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (spring): 27–47.
- Daniel Powers, and Christopher Ellison. 1995. "Interracial Contact and Black Racial Attitudes: The Contact Hypothesis and Selectivity Bias." *Social Forces* 74 (September): 206–26.
- Robert Putnam (with Robert Leonardi and Raffaella Nanetti). 1993. *Making Democracy Work. Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Robert Putnam. 2000. *Bowling Alone: The Collapse and Renewal of American Community*. New York: Simon & Schuster.
- Robert Robinson, and Elton Jackson. 2001. "Is Trust in Others Declining in America? An Age-Period-Cohort Analysis." *Social Science Research* 30 (March): 117–45.
- Richard Rose. 1995. "Russia as an Hour-Glass Society: A Constitution Without Citizens." *East European Constitutional Review* 4 (summer): 34–42.
- Tom W. Smith. 1997. "Factors Relating to Misanthropy in Contemporary American Society." *Social Science Research* 26 (June): 170–96.
- Piotr Sztompka. 1999. *Trust: A Sociological Theory*. Cambridge: Cambridge University Press.
- Eric Uslaner. 2002. *The Moral Foundations of Trust*. New York: Cambridge University Press.
- Eric Uslaner, and Richard Conley. 2003. "Civic Engagement and Particularized Trust: The Ties that Bind People to their Ethnic Communities." *American Politics Research* 31 (4): 331–60.
- Barry R. Weingast. 1998. "Constructing Trust: The Political and Economic Roots of Ethnic and Regional Conflict." In *Institutions and Social Order*, ed. Karol Soltan, Eric M. Uslaner, and Virginia, Haufier. Ann Arbor: University of Michigan Press, 163–200.
- Susan Welch, Lee Sigelman, Timothy Bledsoe, and Michael Combs. 2001. *Race and Place. Race Relations in an American City*. Cambridge: Cambridge University Press.
- Paul Zak and Stephen Knack. 2001. "Trust and Growth." *The Economic Journal* 111 (April): 295–321.

# Military Coercion in Interstate Crises

BRANISLAV L. SLANTCHEV *University of California–San Diego*

**M**ilitary mobilization simultaneously sinks costs, because it must be paid for regardless of the outcome, and ties hands, because it increases the probability of winning should war occur. Existing studies neglect this dualism and cannot explain signaling behavior and tacit bargaining well. I present a formal model that incorporates both functions and shows that many existing conclusions about crisis escalation have to be qualified. Contrary to models with either pure sunk costs or tying-hands signaling, bluffing is possible in equilibrium. General monotonicity results that relate the probability of war to an informed player's expected payoff from fighting do not extend to this environment with its endogenous distribution of power. Peace may involve higher military allocations than war. Rational deterrence models also assume that a commitment either does or does not exist. Extending these, I show how the military instrument can create commitments and investigate the difficulties with communicating them.

**I**n an international crisis, states make demands backed by threats to use force. Although these threats can be explicit in diplomatic communications, they will not generally carry much weight unless substantiated by some show of force—military measures designed to convey the commitment to resort to arms if one's demands are not satisfactorily met. To have an impact, this commitment must be credible; it must be in one's interest to carry out the threat if the opponent refuses to comply. In an environment where states possess private information about their valuations, capabilities, or costs, credibility can be established by actions that a state unwilling to fight would not want, or would not dare, to take. Military moves, such as arms buildups, troop mobilizations, and deployments to the potential zone of operations, can alter incentives in a crisis by changing one's expected payoff from the use of force. These are tacit bargaining moves that can restructure the strategic context thereby creating and possibly signaling one's commitments while undermining those of the opponent. How can states use the military instrument to establish commitments, and how does the nature of the instrument affect their ability to communicate them credibly to their adversaries?

There are two distinct mechanisms for credible signaling. In economic models, information can be transmitted reliably by sinking costs—actors burn money to reveal that they value the disputed issue even more. In contrast, theories of interstate crisis bargaining usually rely on choices that increase the difference between backing down and fighting—actors tie their hands by running higher risks of war to reveal their resolve. The first mechanism involves costs that actors pay regardless of outcome, and the second involves costs that actors pay only if they fail to carry out some threat or promise.

Military actions have *both* cost-sinking and hands-tying effects, and so it is imperative that our theories account for them properly. Focusing only on the cost-sinking role has led scholars to dismiss mobilization as a useful signaling device (Fearon 1997; Jervis 1970; Rector 2003), shifting the focus to mechanisms that have hands-tying effects. Domestic audience costs are the most prominent example of such a signaling mechanism (Fearon 1994) and much work has been done on exploring the role of public commitments.<sup>1</sup> Because open political contestation is a feature of democratic polities, democratic leaders are said to be able to signal their foreign policy preferences better, which in turn provides an explanation of the democratic peace. The model reveals a dynamic of crisis escalation that differs from either pure sunk-cost or hands-tying signaling. Moreover, by demonstrating that it is possible to establish credible commitments with purely military means, the analysis weakens the theoretical argument that democracies are better able to signal their private information.

The model further shows that some of the general monotonicity results from Banks (1990) will not extend to an environment where the probability of victory—and hence the distribution of power—is endogenous to state crisis decisions.<sup>2</sup> Banks finds that the probability of war is increasing in the expected benefits from war of the informed actor. If military mobilization did not influence the probability of winning, then his results would extend to this model as well: actors that value the issue more would have higher expected utilities from war. However, mobilization does influence the probability of winning and, through it, the expected utility of war. Therefore, actors that value the issue more may or may not have higher expected utilities for war, depending on their relative preparedness to wage

Branslav L. Slantchev is Assistant Professor, Department of Political Science, 9500 Gilman Drive, University of California, San Diego, La Jolla, CA 92093-0521 (slantchev@ucsd.edu).

I am especially grateful to Robert Powell for extensive discussions. I thank James Morrow, David Lake, Jaehoon Kim, Joel Watson, Hein Goemans, Randall Stone, and Songying Fang for useful suggestions. Support from the National Science Foundation (Grant SES-0518222) is gratefully acknowledged. Finally, thanks go to Robert Walker and Kris Ramsay for clarifying the Chinese robot issue.

<sup>1</sup> See Smith (1998) on the microfoundations of the audience cost mechanism, and Schultz (2001b) for another critique of its shortcomings.

<sup>2</sup> Banks (1990) establishes results that must be shared by all models with one-sided private information about benefits and costs of war regardless of their specific game-theoretic structure. These generic results turn out to need the additional assumption that the expected payoff from war cannot be manipulated by the actors directly, the very assumption this article questions.



it—the level of which they choose during bargaining. Hence, some standard ideas about crisis escalation that depend on an exogenously fixed distribution of power may need to be modified.

Finally, the analysis illuminates what turns out to be an important shortcoming of existing rational deterrence models. These generally postulate preferences between capitulation and fighting—a resolved type prefers to fight, and an unresolved type prefers to submit—and then explore the consequences of uncertainty about which particular type one is facing. Thus, the models assume that commitments exist and the problem is one of communicating them credibly. I show how the military instrument can *create* commitments and then investigate how this can help with complete information but can hinder the prospects for peace when they have to be communicated under asymmetric information. Unfortunately, whereas mobilization can credibly commit an actor to stand firm, under uncertainty that actor may fail to allocate enough resources to undo his opponent's reciprocal commitment. In this situation, war can become preferable to capitulation for both.

## COERCIVE EFFECTS OF MILITARIZED CRISIS BEHAVIOR

Perhaps the main problem that leaders face in a crisis is credibility: how does a leader persuade an opponent that his threat to use force is genuine? That he would follow up on it should the opponent fail to comply with his demands? The decision to carry out the threat depends on many factors, some or all of which may be unobservable by the opponent. The leader has to communicate enough information to convince her that he is serious. If the opponent believes the message and wants to avoid war, she would be forced to make concessions. However, if there exists a statement that would accomplish this, then all leaders—resolved and unresolved alike—would make it, and hence the opponent would have no reason to believe it. The problem then is to find a statement that only resolved leaders would be willing to make.

Jervis (1970) studies *signals*, which do not change the distribution of power, and *indices*, which are either impossible for the actor to manipulate (and so are inherently credible) or are too costly for an actor to be willing to manipulate. In modern terms, he distinguishes between “cheap talk” and “costly signaling,” even though he prefers to emphasize psychological factors that influence credibility.

It is well known that the possibilities for credible revelation of information when talk is cheap are rather limited and depend crucially on the degree of antagonism between the actors (Crawford and Sobel 1982; Morrow 1994b).<sup>3</sup> Following Schelling (1960), most

studies have explored tacit communication through actions instead of words. Schelling (1966) noted that tactics that reveal willingness to run high risks of war may make threats to use force credible. In general, such willingness results in better expected bargains in crises (Banks 1990), although it does not necessarily mean that the actor willing to run the highest risks would get the best bargain (Powell 1990).

One can think of such tactics in terms of expected benefits from war and expected costs of avoiding it: anything that increases one relative to the other could commit an actor by *tying his hands* at the final stage. Fearon (1994) noted that domestic political audiences can generate costs for leaders who escalate a crisis and then capitulate, creating an environment in which a leader could tie his hands and, thus, signal resolve to foreign adversaries. Even though leaders pay the costs only if they back down, their willingness to risk escalation to a point where each of them would be irrevocably committed to not backing down can reveal their resolve.

This contrasts with another signaling mechanism that relies on *sinking costs*; that is, incurring expenses that do not directly affect the expected payoffs from war and capitulation (Spence 1973). Only actors who value the issue sufficiently would be willing to pay these costs, turning them into a credible revelation of resolve by separating from low-resolve actors through their action. When the last clear chance to avoid war comes, these costs are sunk and cannot affect the decision to attack, hence they cannot work as a commitment device and their function is purely informational.

What is the role of military actions, such as mobilization, in a crisis? Fearon (1994, 579) notes that the “informal literature on international conflict and the causes of war takes it as unproblematic that actions such as mobilization ‘demonstrate resolve,’” and argues that “if mobilization is to convey information and allow learning, it must carry with it some cost or disincentive that affects low-resolve more than high-resolve states.” He then goes on to dismiss the financial costs of mobilization as being insufficient to generate enough disincentive to engage in it and concludes that we should focus on an alternative mechanism—domestic political costs—that has a hands-tying effect.

Although one may quibble with the notion that mobilization is not costly enough, the more important omission is that the argument treats mobilization (and similar militarized crisis activities) as costly actions that are unrelated to the actual use of force. However, one can hardly wage war without preparing for it, and the primary role of mobilization is not to incur costs but rather to prepare for fighting by increasing the chances of victory. But improving one's prospects in fighting increases the value of war relative to peace and can therefore have a hands-tying effect. In fact, it is difficult to conceive of pure sunk costs in this context. Perhaps military exercises away from the potential war zone could qualify as such, but almost anything countries can do in terms of improving defenses or enhancing offensive capability affects the expected payoff from fighting quite apart from the costs incurred in doing it.

<sup>3</sup> Reputational concerns due to continuing interaction with domestic (Guisinger and Smith 2002) or foreign (Sartori 2002) audiences may lend credibility to cheap talk. When both cheap talk and costly messages are available, costly signals can improve the precision of communication (Austen-Smith and Banks 2000).

Even though he does not analyze it, Fearon (1997, n. 27) does recognize this and notes that "insofar as sunk-cost signals are most naturally interpreted as money spent building arms, mobilizing troops, and/or stationing them abroad... the probability of winning a conflict... should increase with the size of the signal."

Underestimating mobilization's role as a commitment device beyond its immediate costliness leads one influential study to conclude that "the financial costs of mobilization rarely seem the principal concern of leaders in a crisis" (Fearon 1994, 580), implying that these costs are insufficient to generate credible revelation of resolve. As I will show, this is true only if mobilization functions solely as a sunk cost; if we consider its hands-tying function, mobilization does acquire crisis bargaining significance.<sup>4</sup> It affects not only signaling behavior of the potential revisionist but also the defensive posture of the status quo power.

Empirically then, it seems that military actions which states take during a crisis—mobilizing troops, dispatching forces—entail costs that are paid regardless of the outcome, and in this sense are sunk; however, they also improve one's expected value of war relative to peace, and in this sense they can tie one's hands. *Militarized coercion* involves actions with these characteristics, but existing theories of interstate crisis bargaining have not analyzed their consequences properly.

In the formal literature, the issue has been almost completely side-stepped in favor of models that incorporate only one of the two functions: the probability of winning is exogenously fixed instead of being determined endogenously by the decisions of the actors.<sup>5</sup> This class of models is nearly exhaustive: very few admit endogenous probability of victory. I am aware of four exceptions. Brito and Intriligator (1985) study resource redistribution as alternative to war under incomplete information but analyze Nash equilibria that may not be sequential (so threats may not be credible) and assume military allocations are made simultaneously (and so one cannot react to the mobilization of the other). Powell (1993) studies the guns versus butter trade-off, but, because he analyzes the complete information case, we cannot use the results to study signaling issues. Kydd's (2000) analysis of bargaining and arms races concentrates on complete information, and the treatment of uncertainty is limited to the special case of two types. Due to the structure of the model, information is revealed at the stage that precedes armament decisions. Consequently, Kydd finds that the informed player's arming choice—that it can potentially use for signaling—is "not really affected by uncertainty; she will arm at whatever level is optimal for her" (238). This is fine for investigating whether arms races can

occur in equilibrium, but constraining for a model that focuses on the potential signaling role of the military instrument. As we shall see, uncertainty does have a significant impact on mobilization levels. Finally, the most closely related approach is that of Morrow (1994a), who models the effect of an alliance as having a dual role: increasing the expected value of war and decreasing the value of the status quo. However, the costs of alliance are not truly sunk because the player does not pay them if it capitulates. As a result the solutions differ significantly from the ones I present here.<sup>6</sup>

In other words, nearly all existing models cannot seriously investigate the impact of military moves in crisis situations because they ignore the hands-tying effect they may have. This is an important shortcoming because, in these models, the probability of winning determines the expected payoff from war, which in turn determines the credibility of threats and, hence, the actor's ability to obtain better bargains. As Banks (1990) demonstrates, the higher the informed actor's expected payoff from war, the higher his payoff from setting the dispute peacefully, and the higher the probability of war in equilibrium. All crisis bargaining models that treat the probability of winning as exogenous would produce this dynamic. However, as I argued, this crucial variable that essentially generates optimal behavior in crisis bargaining models should be part of the process that depends on it. If deliberate actions influence its value, which in turn affects the informational content of these actions, how are we to interpret mobilization decisions? To what extent are costly military actions useful in communicating in crisis: do they make crises more or less stable? What levels of military mobilizations should we expect and what is the price of peace in terms of maintenance of military establishment by defenders?

To answer such questions, the model must have the following features: (a) both actors should be able to choose the level of military mobilization as means of tacit communication, (b) an actor's mobilization should be costly but should increase its probability of winning if war breaks out, (c) mobilization may not necessarily increase the expected utility from war (even though it makes victory more likely, a positive impact, its cost enters negatively), (d) at least one of the actors should be uncertain about the valuation of the other, and (e) actors should be able to make their deliberate attack decisions in light of the information provided by the mobilization levels. Consequently, the model I construct in the next section incorporates all of these.

## THE MODEL

Two players,  $S_1$  and  $S_2$ , face a potential dispute over territory valued at  $v_1 \in (0, 1)$  by the status quo power

<sup>4</sup> Rector (2003) analyzes the impact of mobilization on crisis bargaining but only considers it as partial prepayment of war costs. Because it ignores the hands-tying impact, the study concludes that mobilization has no signaling effect.

<sup>5</sup> This also holds for models where the power distribution changes independently of the choices of the actors, as in Powell (1999, chap. 4) and Slantchev (2003).

<sup>6</sup> Although the economic analysis of contests is closely related to the optimal resource allocation issue (Hirshleifer 1988), the contest models do not allow actors to make their war initiation decisions in light of the new information furnished by the mobilization levels, an important feature of sequential crisis bargaining (Morrow 1989).

$S_1$ , who is currently in possession of it. Although this valuation is common knowledge, the potential revisionist  $S_2$ 's valuation is private information.<sup>7</sup>  $S_1$  believes that  $v_2$  is distributed on the interval  $[0, 1]$  according to the cumulative distribution function  $F$  with continuous strictly positive density  $f$ , and this belief is common knowledge.

Initially,  $S_1$  decides on his military allocation level,  $m_1 \geq 0$ . Choosing  $m_1 = 0$  is equivalent to relinquishing the claim to the territory and ending the game with payoffs  $(0, v_2)$ . Otherwise, the amount  $m_1 > 0$  is invested in possible defense. The costs of mobilization are sunk and incurred immediately. After observing his choice,  $S_2$  either decides to live with the status quo or makes a demand for the territory by starting a crisis.  $S_2$  can escalate by choosing a level of mobilization,  $m_2 > 0$ , or can opt for the status quo with  $m_2 = 0$ , ending the game with the payoffs  $(v_1 - m_1, 0)$ . After observing  $S_2$ 's level of mobilization,  $S_1$  can capitulate, ending the game with payoffs  $(-m_1, v_2 - m_2)$ ; preemptively attack, ending the game with war; or resist, relinquishing the final choice to  $S_2$ . If he resists,  $S_2$  decides whether to capitulate, ending the game with payoffs  $(v_1 - m_1, -m_2)$ , or attack, ending the game with war.

If war occurs, each player suffers the cost of fighting,  $c_i \in (0, 1)$ . Victory in war is determined by the amount of resources mobilized by the players and the military technology. Defeat means the opponent obtains the territory. The probability that player  $i$  prevails is  $\lambda m_i / (\lambda m_i + m_{-i})$ , where  $\lambda > 0$  measures the offense–defense balance.<sup>8</sup> If  $\lambda = 1$ , then there are no advantages to striking first. If  $\lambda > 1$ , then offense dominates and, for any given allocation  $(m_1, m_2)$ , the probability of prevailing by striking first is strictly larger than the probability of prevailing if attacked. Conversely, if  $\lambda < 1$ , then defense dominates, and for any given allocation it is better to wait for an attack instead of striking first. If  $i$  attacks first, the expected payoff from war is  $W_i^a(m_1, m_2) = \lambda m_i v_i / (\lambda m_i + m_{-i}) - c_i - m_i$ , and, if  $i$  is attacked, it is  $W_i^d(m_1, m_2) = m_i v_i / (m_i + \lambda m_{-i}) - c_i - m_i$ . It is easy to show that  $\lambda < 1 \Leftrightarrow W_i^d > W_i^a$ . If defense dominates, then the expected value of war is higher when one is attacked than when one attacks first.<sup>9</sup> For the rest of this paper, assume  $\lambda < 1$ . The central claims do not change when  $\lambda > 1$ , but the statement

of the results is quite a bit more involved (Slantchev 2004a).

The solution concept is perfect Bayesian equilibrium (or simply “equilibrium”), which requires that strategies are sequentially rational given the beliefs, and that beliefs are consistent with the strategies, and derived from Bayes rule whenever possible (Fudenberg and Tirole 1991). The model incorporates the empirically motivated features I identified in the preceding section. It is complicated by the continuum of types and actions, and so it trades an ultimatum “bargaining” protocol for rich mobilization possibilities in letting both actors choose the level of forceful persuasion.

## THE MOBILIZATION OF THE REVISIONIST STATE

It will be helpful to analyze the signaling game beginning with  $S_2$ 's allocation decision given some allocation  $m_1 > 0$ . In any equilibrium, the strategies would have to form an equilibrium in this continuation game, and since  $S_1$  is uninformed, his initial decision reduces to choosing (through his allocation) the equilibrium that yields the highest expected payoff.

By subgame perfection,  $S_2$  would attack at her final decision node if, and only if, her expected payoff from war is at least as good as capitulating:  $W_2^a(m_1, m_2) \geq -m_2$ . That is,  $v_2 \geq c_2 + c_2 m_1 / (\lambda m_2) \equiv \gamma(m_1, m_2) > 0$ , where  $\gamma(m_1, m_2)$  is the highest type that would capitulate if resisted at the allocation level  $(m_1, m_2)$ . All types  $v_2 < \gamma(m_1, m_2)$  capitulate, and all types  $v_2 \geq \gamma(m_1, m_2)$  attack when resisted. Note that  $\gamma(m_1, m_2) > 0$  implies that the lowest-valuation types never attack even if they are certain to win. For any posterior belief characterized by the distribution function  $G(\gamma(m_1, m_2))$  that  $S_1$  may hold, resisting at the allocation  $(m_1, m_2)$  yields  $S_1$  the following expected payoff:  $R_1(m_1, m_2) = G(\gamma)(v_1 - m_1) + (1 - G(\gamma))W_1^d(m_1, m_2)$ . If  $S_1$  attacks preemptively, he would get  $W_1^a(m_1, m_2)$ . Since  $W_1^d(m_1, m_2) < v_1 - m_1$ , it follows that  $\lambda < 1 \Rightarrow W_1^d(m_1, m_2) < R_1(m_1, m_2)$  regardless of  $S_1$ 's posterior belief. Therefore, if defense dominates, then in equilibrium  $S_1$  never preempts; he either capitulates or resists.

Suppose that  $S_1$  capitulates for sure if he observes an allocation  $\bar{m}_2(m_1)$ . There can be at most one such *assured compellence level* in equilibrium. To see that, suppose that there were more than one. But then all  $S_2$  types who allocate the higher level can profit by switching to the lower one. Obviously,  $\bar{m}_2(m_1)$  is an upper bound on any equilibrium allocation by  $S_2$ . Furthermore  $S_2$  would never mobilize  $m_2 \geq 1$  in any equilibrium. This is because the best possible payoff she can ever hope to obtain is  $v_2 - m_2$  if  $S_1$  capitulates, and this is nonpositive for any  $m_2 \geq 1$ , for all  $v_2 \leq 1$ .

All of this suggests that  $S_2$ 's equilibrium behavior would be determined by the relationship among the payoffs she can obtain from optimal offensive war, assured compellence, and capitulation. That is,  $S_2$ 's strategy can be characterized by a series of cut-points that divide her types into subsets who behave the same way. To this end, I now derive these cut-points and

<sup>7</sup> Since  $S_1$  has the territory, it is natural to assume that his valuation is known to everyone. The labels “status quo power” and “potential revisionist” identify which actor would be in possession of the territory if a crisis does not occur. This has nothing to do with the degree of satisfaction with the status quo that determines these labels in classical realism. For ease of exposition, I refer to  $S_1$  as a “he” and  $S_2$  as a “she.”

<sup>8</sup> The ratio form of the contest success function is undefined at  $m_1 = m_2 = 0$ , but since the game ends with  $m_1 = 0$ , how we define it is immaterial.

<sup>9</sup> This offense–defense balance depends on military technology and differs from the ease of conquest concept that goes under the same name in offense–defense theory (Jervis 1978, Quester 1977). According to that theory, “offense–defense balance” refers to whether it is easier to take a territory than to defend it. Because the territory belongs to  $S_1$  in this model, a defensive advantage means that  $S_1$  would defend it more easily given the same distribution of power than  $S_2$  could acquire.

then show that only two configurations can occur in equilibrium.

Let  $\beta(m_1)$  denote the type that is indifferent between optimal war and assured compellence; that is,  $W_2^a(m_1, m_2^*(m_1, \beta(m_1))) = \beta(m_1) - \bar{m}_2(m_1)$ , where  $m_2^*(m_1, v_2) = \sqrt{m_1 v_2 / \lambda} - m_1 / \lambda > 0$  is the optimal allocation by type  $v_2$  if she expects to fight for sure some  $m_1$ . That is,  $m_2^*(m_1, v_2)$  maximizes  $W_2^a(m_1, m_2(v_2))$  subject to the constraint that  $m_2^* > 0$ . Substituting and solving for  $\beta(m_1)$  yields  $\beta(m_1) = (m_1 + \lambda(\bar{m}_2(m_1) - c_2))^2 / (4\lambda m_1)$ . The following lemma establishes the  $S_2$ 's preference between optimal war and assured compellence (all proofs are in the Appendix).

**LEMMA 1.** *All  $v_2 > \beta(m_1)$  strictly prefer assured compellence to optimal war, and all  $v_2 \leq \beta(m_1)$  prefer the opposite.*

Let  $\alpha(m_1)$  denote the type that is indifferent between capitulation and assured compellence at  $\bar{m}_2(m_1)$ ; that is,  $\alpha(m_1) - \bar{m}_2(m_1) = 0$ . Since the payoff from assured compellence strictly increases in type, all  $v_2 < \alpha(m_1)$  prefer capitulation to assured compellence, and all  $v_2 \geq \alpha(m_1)$  prefer assured compellence to capitulation.

Let  $\delta(m_1)$  denote the type that is indifferent between capitulation and optimal war. That is,  $W_2^a(m_1, \delta(m_1)) = 0$ , which yields  $\delta(m_1) = c_2 + 2\sqrt{c_2 m_1 / \lambda} + m_1 / \lambda$ . Since the payoff from optimal war is strictly increasing in type, all  $v_2 < \delta(m_1)$  prefer capitulation to optimal war, and all  $v_2 \geq \delta(m_1)$  prefer optimal war to capitulation.

I now establish the possible configurations of these cut-points. With slight abuse of notation, I suppress their explicit dependence on  $m_1$ .

**LEMMA 2.** *If  $\alpha \leq \delta$ , then all  $v_2 < \alpha$  capitulate and all  $v_2 \geq \alpha$  mobilize at the compellence level  $\bar{m}_2(m_1)$  in equilibrium, provided  $\bar{m}_2(m_1)$  is feasible.*

Lemma 2 shows that, when  $\delta \geq \alpha$ , optimal behavior can take only one form if  $\bar{m}_2(m_1)$  is feasible.<sup>10</sup> Hence, we need not worry about the location of  $\beta$ . The following lemma establishes that only one configuration remains for the other case.

**LEMMA 3.** *If  $\delta < \alpha$ , then  $\alpha < \beta$ .*

These lemmata imply that we should look for solutions for just two cut-point configurations:  $\alpha \leq \delta$  and  $\delta < \alpha < \beta$ . Optimal behavior depends on the relationship between these points and  $S_2$ 's highest valuation (unity).

## Assured Compellence

In an assured compellence equilibrium, all types of  $S_2$  that mobilize do so at a level just enough to make  $S_1$  capitulate with certainty. Intuitively, if  $S_1$  has mobilized at a low level, it is relatively easy for  $S_2$  to countermobilize such that  $S_1$ 's payoff from war becomes sufficiently low. This undermines  $S_1$ 's incentive to resist even if there still exists a chance that  $S_2$  is bluffing. Despite  $S_1$ 's certain capitulation, not all low-valuation types will be willing to bluff because of the inherent costliness of mobilization. Hence, we shall look for an equilibrium in which all low-valuation types capitulate, and the rest allocate the assured compellence level.  $S_1$  resists all allocations smaller than this level (because only low-valuation types that would capitulate if resisted would fail to allocate the higher level) and capitulates otherwise.

Suppose  $\alpha \leq \delta$  and  $\alpha < 1$ . By Lemma 2,  $S_2$ 's optimal strategy must take the following form: all  $v_2 < \alpha$  capitulate immediately, and all  $v_2 \geq \alpha$  mobilize at the compellence level  $\bar{m}_2$ . By definition,  $\alpha - \bar{m}_2 = 0$ , and therefore  $\alpha = \bar{m}_2$ . If  $\bar{m}_2 < 1$ , then the assured compellence level is feasible because there exists a type of  $S_2$  that could choose to allocate  $\bar{m}_2$  optimally, and so  $S_1$  is *potentially compellable*. Otherwise, he is *uncompellable*.

Subgame perfection implies that, if  $\alpha \leq \gamma(m_1, \bar{m}_2)$ , all types  $v_2 < \gamma(m_1, \bar{m}_2)$  capitulate if resisted (bluffers) and all  $v_2 \geq \gamma(m_1, \bar{m}_2)$  fight if resisted (genuine challengers). If  $\alpha > \gamma(m_1, \bar{m}_2)$ , only genuine challengers mobilize in equilibrium. Given  $S_1$ 's prior belief  $F(\cdot)$ , his posterior belief that  $S_2$  would capitulate when resisted conditional on  $\bar{m}_2$  is  $G(\gamma(m_1, \bar{m}_2)) = (F(\gamma(m_1, \bar{m}_2)) - F(\bar{m}_2)) / (F(1) - F(\bar{m}_2))$  if  $\bar{m}_2 \leq \gamma(m_1, \bar{m}_2)$ , and 0 otherwise.  $S_1$  capitulates whenever  $R_1(m_1, m_2) \leq -m_1$ . Because  $R_1$  is strictly decreasing in  $m_2$  and because excess mobilization by  $S_2$  is pointless if  $S_1$  is sure to capitulate, it follows that in equilibrium  $\bar{m}_2$  must solve  $R_1(m_1, \bar{m}_2) = -m_1$ , or

$$G(\gamma(m_1, \bar{m}_2))v_1 + [1 - G(\gamma(m_1, \bar{m}_2))] \times \left[ \frac{m_1 v_1}{m_1 + \lambda \bar{m}_2} - c_1 \right] = 0. \quad (1)$$

Let  $\bar{m}_2$  be the unique solution to equation (1).<sup>11</sup>

**Proposition 1 (Assured Compellence).** *Fix some  $m_1$ . If and only if  $\alpha < \delta$  and  $\alpha < 1$ , the following strategies constitute the unique equilibrium in the continuation game: all  $v_2 < \alpha$  capitulate, and all  $v_2 \geq \alpha$  allocate  $\bar{m}_2$ ; if resisted, all  $v_2 < \gamma$  capitulate, and all  $v_2 \geq \gamma$  attack.  $S_1$  resists after any  $m_2 < \bar{m}_2$  and capitulates after any  $m_2 \geq \bar{m}_2$ .*

<sup>10</sup> Technically, any  $m_2 > 0$  is feasible because there is no budget constraint. However, since  $S_2$  would never spend more than her highest possible valuation in equilibrium, this valuation functions as an effective constraint. The results remain unchanged if we allow for an arbitrary upper bound on valuations except we would have to restate the theorems in terms of that bound.

<sup>11</sup> To see that equation (1) has a unique solution, let  $\hat{m}_2 \equiv 1/2[c_2 + \sqrt{c_2(c_2 + 4m_1/\lambda)}]$  and note that  $m_2 \leq \hat{m}_2 \Leftrightarrow m_2 \leq \gamma(m_1, m_2)$ . This implies that for all  $m_2 \geq \hat{m}_2$ ,  $G(\gamma(m_1, m_2)) = 0$ . Equation (1) is strictly decreasing in  $\bar{m}_2$ , and for all  $\bar{m}_2 \geq \hat{m}_2$  it reduces to  $m_1 v_1 / (m_1 + \lambda \bar{m}_2) - c_1$ , which itself converges to  $-c_1 < 0$  as  $\bar{m}_2 \rightarrow \infty$ . Because the expression is continuous in  $\bar{m}_2 > 0$ , it follows that equation (1) has a unique solution.

There is no risk of war in this equilibrium because whenever a positive mobilization occurs the crisis is resolved with  $S_1$ 's capitulation. If  $S_1$  allocates too little to defense, he can expect that  $S_2$  will challenge him with strictly positive probability and he will capitulate. This does not necessarily mean that  $S_1$  immediately gives up the territory in equilibrium: as long as the probability of a challenge is not too high,  $S_1$  is still better off spending on defense and taking his chances that  $S_2$ 's valuation would not be high enough to challenge him. This equilibrium involves bluffing whenever  $\bar{m}_2 < \hat{m}_2$ , which cannot be eliminated with an appeal to any of the refinements like the intuitive criterion (Cho and Kreps 1987), universal divinity (Banks and Sobel 1987), or perfect sequentiality (Grossman and Perry 1986). Although nongenuine challengers may be present, their bluff is never called.

### Risk of War

When  $S_1$ 's mobilization level increases,  $S_2$ 's countermobilization required to achieve assured compellence increases as well. As ensuring that outcome gets costlier, risking optimal war becomes more attractive. In particular, if the type who is indifferent between war and capitulation has a lower valuation than the type who is indifferent between assured compellence and capitulation, all intermediate-valuation types would rather fight than ensure the exceedingly costly capitulation by  $S_1$  or give up themselves. Increasing  $m_1$  even further eliminates all possibility that some type would be willing to attempt compellence, reducing  $S_2$ 's choice to capitulation or optimal war.

Turning to the formal statement of this result, suppose  $\delta < \alpha$ . By Lemma 3, only one possible configuration exists:  $\delta < \alpha < \beta$ . Since all  $v_2 > \delta$  prefer optimal war to capitulation, all challenges in this equilibrium are genuine, and  $G = 0$  simplifies equation (1) yielding an analytic solution to the compellence level  $\alpha = \bar{m}_2 = m_1(v_1 - c_1)/(\lambda c_1)$ . This is also the solution to equation (1) for the assured compellence equilibrium when  $\bar{m}_2 > \gamma(m_1, \bar{m}_2)$ . Substituting for  $\bar{m}_2$  yields  $\beta = (1/4\lambda m_1)(\lambda c_2 - m_1 v_1/c_1)^2$ .

**Proposition 2 (Risk of War).** *Fix some  $m_1$ . If, and only if,  $\delta \leq \alpha$  and  $\delta < 1$ , the following strategies constitute the unique equilibrium of the continuation game: all  $v_2 < \delta$  capitulate, all  $v_2 \in [\delta, \beta)$  allocate  $m_2^*(m_1, v_2)$ , and all  $v_2 > \beta$  allocate  $\bar{m}_2$ ; if resisted, all  $v_2 < \gamma$  capitulate, and all  $v_2 \geq \gamma$  attack.  $S_1$  resists after any  $m_2 < \bar{m}_2$  and capitulates after any  $m_2 \geq \bar{m}_2$ .*

All challengers in this equilibrium are genuine. The outcome depends on whether  $S_1$  is potentially compellable and whether there exists a type of  $S_2$  that is willing to allocate at the assured compellence level.

If  $\alpha < \beta < 1$ , the ex ante probability of war is  $\Pr(\delta \leq v_2 < \beta) = F(\beta) - F(\delta) < 1$ . If  $S_2$  has a high enough valuation  $v_2 > \beta$ , then she would allocate at the assured compellence level and  $S_1$  would capitulate. The most dangerous revisionists are the midrange valuation types  $v_2 \in [\delta, \beta)$ , the ones who do not value the issue

sufficiently to spend the amount necessary to ensure  $S_1$ 's peaceful concession. Even though  $S_1$  is potentially compellable, these types are unwilling to do it, and they go to war choosing their optimal attack allocation. It is worth noting that because they separate fully by their optimal allocation,  $S_1$  infers their type with certainty and knows that resistance would mean war because all challenges are genuine. If the revisionist happens to be of such a type, then war occurs with complete information following her mobilization.

If  $\alpha < 1 \leq \beta$ , then even though  $S_1$  is potentially compellable, no type is willing to do it, and war is certain conditional on a challenge. Because  $\delta$  is strictly increasing in  $m_1$ , it follows that higher allocations by  $S_1$  never increase the risk of war. (If  $F$  has continuous and strictly positive density, then increasing  $m_1$  strictly decreases the risk of war.) Unlike the previous case, the most dangerous revisionists here are always the ones with higher valuations  $v_2 \geq \delta$  because they cannot be deterred from challenging.  $S_1$  infers the revisionist's type with certainty and war occurs with complete information conditional on a mobilization by  $S_2$ . I shall refer to this as the risk of war, type 1 equilibrium.

Finally, if  $1 \leq \alpha$ , then  $S_1$  becomes uncompellable and  $S_2$ 's choice reduces to capitulation or optimal attack. From  $S_1$ 's ex ante perspective, the situation is identical to the preceding case where no type was willing to compel him, except that now no type is able to do so. Higher allocations by  $S_1$  never increase the risk of war in this case, and the most dangerous types are the high-valuation ones. I shall refer to this as the risk of war, type 2 equilibrium.

### Assured Deterrence

Finally, if  $S_1$  mobilizes at a very high level, then he can become uncompellable and no types would be willing to challenge him given that he is certain to resist. In other words,  $S_1$  can achieve assured deterrence. This can happen when there is no type that is willing to fight even an optimal war, and when the assured compellence level is not feasible. The following proposition states the necessary and sufficient conditions for this equilibrium.

**Proposition 3 (Assured Deterrence).** *Fix some  $m_1$ . If, and only if,  $\alpha \geq 1$  and  $\delta \geq 1$ , the following strategies constitute the unique equilibrium of the continuation game: all  $v_2$  capitulate; if resisted, all  $v_2 < \gamma$  capitulate, and all  $v_2 \geq \gamma$  attack.  $S_1$  resists all allocations.*

The probability of war is zero and the outcome is capitulation by  $S_2$ . To understand the conditions, note that, when  $\alpha > \delta$  (as it would be in transitioning from the risk of war equilibrium),  $\delta \geq 1$  is sufficient. However, it is possible to transition from the assured compellence equilibrium directly. To see this, note that, since  $\alpha < 1$  and  $\alpha < \delta$  are necessary and sufficient for that equilibrium, then  $\alpha \geq 1$  is sufficient for it to fail to exist, and  $\alpha < \delta$  further implies  $\delta > 1$ , and so it is also sufficient for deterrence to exist as long as  $\alpha < \delta$ . In other words,

the configurations  $1 \leq \delta < \alpha$  and  $1 \leq \alpha < \delta$  both result in deterrence.

## THE DEFENSE OF THE STATUS QUO STATE

Collectively, the three mutually exclusive equilibria exhaust all possible configurations of the cut-points and, therefore, provide the solution for the continuation game for any set of the exogenous parameters and any  $m_1 > 0$ . I now turn to  $S_1$ 's initial mobilization decision. Since  $S_1$  is the uninformed actor, his choice boils down to selecting which type of equilibrium will occur in the continuation game. It is not possible to derive an analytic solution to this problem because of the nonlinearities involved in the optimization at the second stage. Still, because we can generally establish the order in which the continuation game equilibria occur as a function of  $m_1$ , we can say what type of choices  $S_1$  will face if he increases his mobilization level. With the help of computer simulations, we can derive precise predictions for interesting ranges of the exogenous variables too.

The compellence equilibrium always exists regardless of the values of the exogenous parameters because, for  $m_1$  small enough, the necessary and sufficient condition from Proposition 1 are satisfied. What happens once  $m_1$  begins to increase? As the derivations in the previous section suggest, two cases are possible. First, as  $m_1$  increases, the conditions for deterrence can be satisfied, and the continuation game has only two possible solutions, both involving peace. Second, as  $m_1$  increases, the existence conditions can successively satisfy the risk of war and deterrence equilibria.

To see how  $S_1$  would choose his initial mobilization, if any, we must consider his expected payoffs in each of the possible continuation game equilibria. To conduct comparative statics simulations and analyses, I impose the additional assumption that  $F$  is the uniform distribution. This also allows me to reduce the expected payoffs for  $S_1$  to manageable expressions.

In the compellence equilibrium,  $S_1$  obtains the prize with probability  $\Pr(v_2 \leq \alpha) = \alpha$  by the distributional assumption and concedes it without fighting with complementary probability. His expected payoff is  $EU_1^{\text{COMPEL}}(m_1) = \alpha v_1 - m_1$ . In the risk of war equilibrium,  $S_1$  obtains the prize with probability  $\Pr(v_2 \leq \delta) = \delta$ , fights a war with probability  $\Pr(\delta < v_2 \leq \beta) = \beta - \delta$ , and concedes the prize with probability  $\Pr(v_2 > \beta) = 1 - \Pr(v_2 \leq \beta) = 1 - \beta$ . His expected payoff is

$$\begin{aligned} EU_1^{\text{RISK}}(m_1) &= \delta(v_1 - m_1) + \int_{\delta}^{\beta} W_1^d(m_1, m_2^*(x)) f(x) dx - (1 - \beta)m_1 \\ &= \left[ \delta + 2\sqrt{\frac{m_1}{\lambda}}(\sqrt{\beta} - \sqrt{\delta}) \right] v_1 - (\beta - \delta)c_1 - m_1, \end{aligned}$$

where we used  $W_1^d(m_1, m_2^*(v_2)) = v_1/v_2\sqrt{m_1 v_2/\lambda} - c_1 - m_1$ . Finally, in the deterrence equilibrium,  $S_1$ 's payoff

is:  $EU_1^{\text{DETER}}(m_1) = v_1 - m_1$ . In equilibrium there can be only one assured deterrence allocation level by  $S_1$  because, if there were two, then  $S_1$  could profitably deviate to the lower one.

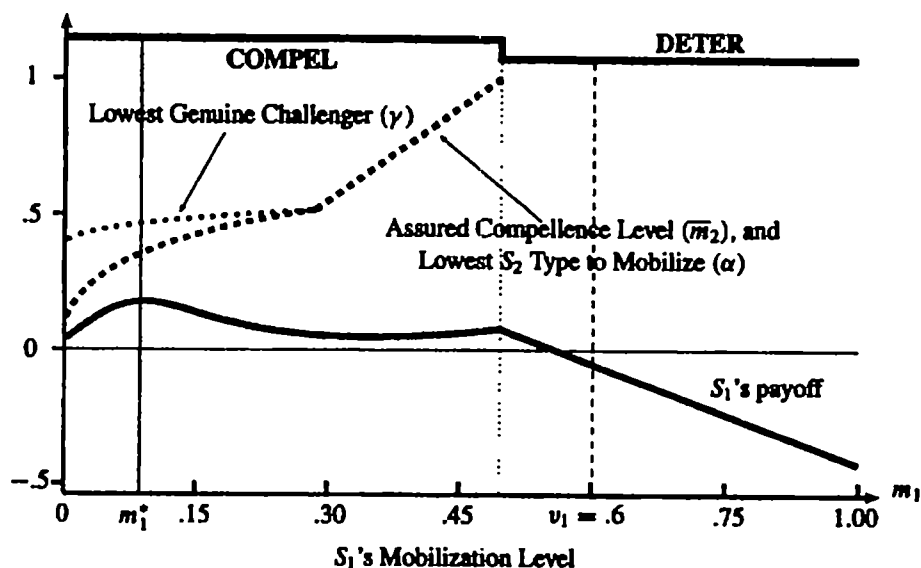
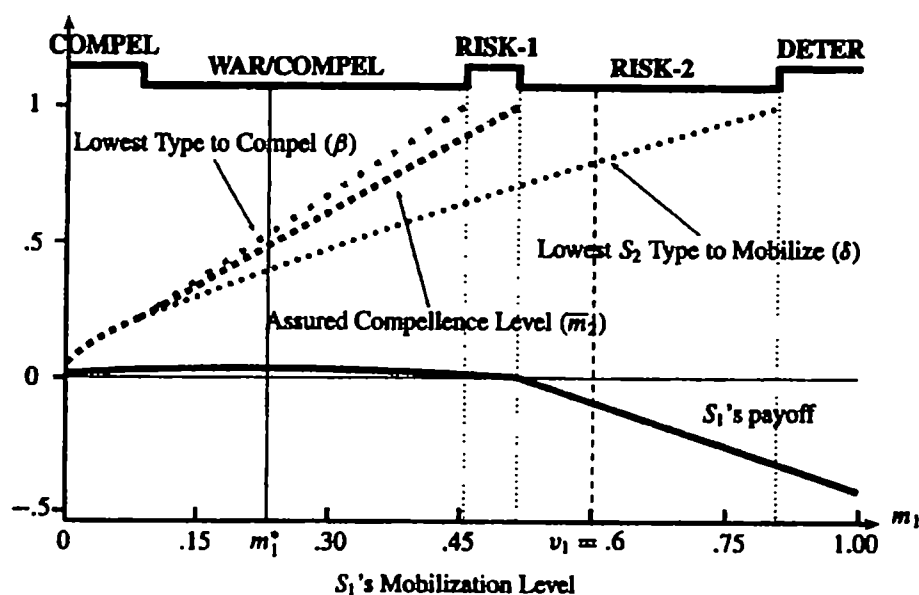
I now provide two numerical examples that will facilitate the substantive discussion. Assume the uniform distribution for  $S_2$ 's valuations, and set the parameters  $v_1 = 0.6$ ,  $c_1 = 0.2$ , and  $\lambda = 0.99$ . In the simulation in Figure 1(a),  $S_2$ 's costs of fighting are high,  $c_2 = 0.35$ , and in the simulation in Figure 1(b), her costs of fighting are low,  $c_2 = 0.01$ . The solid line shows the range of values for  $m_1$  for which the various equilibria exist. The dotted vertical line shows  $S_1$ 's valuation for reference, and the solid vertical line shows  $S_1$ 's equilibrium mobilization level.

In the first example, the equilibrium outcome is peace: one of the actors will capitulate.  $S_1$  mobilizes  $m_1^* = 0.07$  and takes his chances that  $S_2$  may be a high-valuation type that would compel him to capitulate. The assured compellence level is  $\bar{m}_2 = \alpha = 0.33$ . The probability that  $S_1$ 's low mobilization level would be able to deter  $S_2$  is  $\Pr(v_2 < \alpha) = 33\%$ , and so the risk of having to concede is 67%. All types  $v_2 < \alpha$  quit and  $S_1$  gets to keep the territory. On the other hand, all types  $v_2 \geq \alpha$  allocate  $\bar{m}_2$ , after which  $S_1$  relinquishes the territory without a fight.

In the second example, the outcome can be either capitulation by one of the actors or war.  $S_1$ 's optimal mobilization increases to  $m_1^* = 0.25$ . What follows depends on just how high the challenger's valuation is. If it is  $v_2 < \delta = 0.36$ , then  $S_2$  would be deterred from mobilizing, and the outcome would be peace. If it is  $v_2 \geq \beta = 0.55$ , then  $S_2$  would mobilize at the assured compellence level  $\bar{m}_2 = \alpha = 0.50$ ,  $S_1$  would capitulate, and the outcome would be peace again. However, if  $v_2 \in [0.36, 0.55]$ , then  $S_2$  would allocate her optimal fighting level  $m_2^*(v_2) < 0.50$ , and the outcome would be war. The ex ante probability of war is 19%, but conditional on  $S_2$ 's mobilization it is 30%, with war being certain if  $S_2$ 's mobilization level is less than  $\bar{m}_2$ .

$S_1$ 's expected payoff in this equilibrium is 0.02, which is much less than the 0.13 he would have expected in the previous example. This is not surprising, because as  $S_2$ 's costs of fighting decrease, so does  $S_1$ 's equilibrium payoff: to wit, his opponent is able to extract a better deal because going to war is not as painful, and so the threat to do it is much more credible.

These dynamics clearly demonstrate that establishing a credible commitment by tying one's hands can avoid war only if it also makes fighting sufficiently unpleasant to the opponent. A credible threat to fight cannot buy peace by itself, and a perfect commitment can virtually guarantee war if the opponent's valuation is misjudged. It is worth noting that *crises that are peacefully resolved may involve higher military allocations than those that end in war*: either  $S_1$  mobilizes a large enough force to deter  $S_2$ , or  $S_2$  mobilizes a large enough force to compel  $S_1$ . These allocations are higher than the optimal war allocations that either state would make if they expect to fight for sure. In other words, arms buildups are not necessarily destabilizing in a

FIGURE 1. Examples of Equilibria as Function of  $S_1$ 's Mobilization(a) Assured Compellence Equilibrium with Bluffing, High-Cost  $S_2$ (b) Risk of War Equilibrium, Low-Cost  $S_2$ 

crisis. In fact, they appear positively related to peace when it comes to threatening the use of force.

## DISCUSSION

Fearon (1997) nicely brackets the analysis presented here. He analyzes the two polar mechanisms for signaling interests: through actions that involve sunk costs only and actions that tie hands only. My model essentially encompasses everything in between—that is, actions that both tie hands and sink costs—and so it is worth comparing the results.

## Bluffing with Implicit Threats

The most obvious difference that is of great substantive interest is that actions involving each mechanism separately result in equilibria where bluffing is not possible.<sup>12</sup> As it turns out, this result is unstable.

<sup>12</sup> That is, no equilibria that survive the Intuitive Criterion (Cho and Kreps 1987) involve bluffing. Fearon (1997, 82, n. 27) notes that it is unrealistic to assume that “sunk-cost signals have no military impact” and conjectures that the strong no-bluffing result would obtain even when we relax that assumption.



In Fearon's hands-tying model, bluffing cannot occur because actors with high valuations can generate arbitrarily high costs for backing down, which they never have to pay because their opponent would submit. Maximizing the payoff of high-valuation types reduces to maximizing the probability of capitulation by the opponent. This does not work in a model where hands-tying is inherently costly because now maximizing the probability of capitulation by the opponent must be balanced against its costs, which may put a cap on worthwhile mobilization levels, and that in turn can induce lower-valuation types to bluff because it makes it affordable. In addition to its costliness and impact on one's own war payoff, an actor's mobilization also affects the expected war payoff of its opponent. This separates mobilization from the audience-cost models where one's actions have no direct bearing on the opponent's payoffs. In other words, the actors' ability to generate high signals is constrained both by the costliness of the military instrument and by the actions of their opponent.

Take, for example, the assured compellence equilibrium in Figure 1(a). There are bluffers here: all  $v_2 \in [\alpha, \gamma] = [.33, .42]$  would not attack should  $S_1$  decide to resist. The ex ante probability of a bluffer is  $\Pr(\alpha \leq v_2 < \lambda) = 9\%$ , which increases to 13% after  $S_2$  mobilizes. However, even though  $S_1$  is now far more likely to be facing a bluffer, he is also far more likely to be facing a genuine challenger (87% versus an initial 58%), and so he chooses not to resist. The small mobilization has successfully screened out low-valuation types and  $S_1$  is unwilling to run a risk of war at this stage given how much  $S_2$ 's mobilization has reduced his payoff from war. Note that  $S_1$  could have eliminated all bluffers if he wished to do so by allocating approximately  $m_1 = 0.28$  (this is where  $\gamma = \alpha$ ), but doing so is not optimal because of the costs involved. Hence, not only is bluffing possible in equilibrium but  $S_1$  would not necessarily attempt to weed out such challengers. Further,  $S_2$ 's countermobilization has essentially untied  $S_1$ 's hands by lowering his expected payoff from war to the point where capitulation is preferable.

On the other hand, bluffing is impossible in equilibria that involve genuine risk of war. Consider Figure 1(b): there can be no bluffing here, for a bluffer would have to mobilize at the assured compellence level—otherwise she would be forced to back down when  $S_1$  resists and suffer the costs of mobilization—and this level is too high given  $S_1$ 's initial mobilization.

Hence, bluffing is possible only in equilibria that do not involve much revelation of information and involve no danger of war. This corresponds to results of Brito and Intriligator (1985), who also find that in the pooling (no signaling) equilibrium bluffing is possible but the probability of war is zero. Preventing bluffing involves precommitment to a positive probability of war, and the willingness to run this risk does transmit information.

The model reveals a subtle distinction in the conditions that permit bluffing. Bluffing is only optimal when  $S_1$  is expected to capitulate, but his willingness to do

so depends on how likely  $S_2$  is to fight, which in turn depends on  $S_2$ 's costs of fighting and  $S_1$ 's mobilization level. *Paradoxically, bluffing by  $S_2$  is possible only when her costs of fighting are relatively high (she is "weak").* The reason is the effect this has on  $S_1$ 's decision: because  $S_2$  is weak, and therefore not very likely to be willing to mobilize at a high level,  $S_1$  reduces his own costly allocation and thereby exposes himself to the possibility of having to concede. *It is this low mobilization that makes bluffing an option: one must choose to expose oneself to bluffing. It is always possible to eliminate that possibility by making it too dangerous a tactic.* When  $S_2$  has relatively low costs of fighting (she is "strong"),  $S_1$  knows that low mobilization would virtually ensure his capitulation, and so he ups the ante, eliminating bluffing possibilities in the process. Essentially, bluffing becomes too expensive even if it is certain to succeed. For this result to obtain, mobilization must both be inherently costly and increase probability of victory.

Fearon buttresses his no-bluffing results by quoting an observation by Brodie (1959, 272), who states that "bluffing, in the sense of deliberately trying to *sound* more determined or bellicose than one actually felt, was by no means as common a phenomenon in diplomacy . . . it tended to be confined to the more *implicit* kinds of threat." I have emphasized the distinction between verbal threats and implicit threats because it is very important. Reputational concerns may eliminate the incentives to bluff with words (Guisinger and Smith 2002; Sartori 2002) but may not work for implicit threats like the ones in this model. As Iklé (1964, 64) observes, "whether or not the threat is a bluff can be decided only after it has been challenged by the opponent's noncompliance." But probing an implicit threat is too dangerous because by its very nature, and unlike words, it influences the expected outcome of war. In equilibrium, these types of bluffs are never called, and hence  $S_2$  is never revealed as having made an incredible threat.<sup>13</sup> As Powell (1990, 60) concludes, "sometimes bluffing works."

Military coercion is a blunt instrument because its intent is not to reveal the precise valuation of the informed party but rather to communicate one's willingness to fight. Although much nuance is possible if actors had in mind the former goal, the latter is, of necessity, rather coarse. That one must resort to tacit bargaining through implicit threats cannot improve matters. Historians have emphasized the difficulty in clarifying "the distinction between warning and intent" (Strachan 2003, 18). Perhaps it is precisely because mobilization has such a crude signaling role, which is hard to disentangle from preparation for war, that mobilization has traditionally been considered very dangerous.

<sup>13</sup> The result of bluffs never being called in equilibrium probably arises from the one-sided incomplete information in the model. If there were uncertainty about  $S_1$ 's valuation as well,  $S_2$  could bluff hoping that  $S_1$  will quit, and because she does not know her opponent's type, she may end up facing one that is prepared to resist.

## Endogenous Distribution of Power

Military coercion has a somewhat peculiar dynamic completely lost to models that ignore the war-fighting implications of military measures. For example, it is now generally accepted that the stronger the actor, the more willing he is to risk war to obtain a better bargain. The risk–return trade-off then resolves itself in higher equilibrium probability of war and a better expected negotiated deal (Banks 1990; Powell 1999).

Generally, a strong actor is one with a large expected war payoff. Valuation of the issue (high), costs of fighting (low), probability of winning (high), and military capabilities (large) can be lumped together to produce an aggregate expected payoff from fighting (high), which in turn defines the actor's type (strong). Potential opponents can then be indexed by their war payoffs, which are taken to be exogenous to the model. Bargaining essentially involves attempts to discern just how much concessions the opponent is prepared to make, and that in turn depends on how much he expects to obtain by fighting. When the distribution of power is fixed, the only way weak types can be discouraged from mimicking the behavior of strong types and demanding too much is for strong types to run a higher risk of war. Mobilization endogenizes the payoff from fighting, and its costliness provides another way to discourage weak types without necessarily running a higher risk of war.<sup>14</sup>

This now means that we need to pay closer attention to the effects of short-term mobilizations because they help determine, at least in part, the expected payoff from war endogenously.<sup>15</sup> The immediate implication is that incentive-compatibility arguments that rely on an exogenous distribution of power may not extend to this context. For example (Banks 1990, 606) argues that “as the expected benefits from war increase, the informed player receives a better negotiated settlement but in addition runs a greater risk of war.” Because  $S_2$ 's types are distinguished by their valuation of the issue in my model (mobilization and war costs are the same for all types), the equivalent statement would contend that higher-valuation types obtain  $S_1$ 's capitulation with higher probability but also run an increased risk of war.

*The model shows that the expected payoff from the crisis does increase in the actors' valuation of the issue, but not necessarily at the cost of a higher risk of war. In other words, the risk–return trade-off does not neces-*

*sarily operate in this context, where the relevant trade-off is between signaling cost and expected return.* To see that, consider the risk of war equilibrium. All low-valuation types capitulate immediately and so face zero probability of fighting. All midvaluation types mobilize their optimal fighting allocations, and the probability of war jumps to one. On the other hand, high-valuation types manage to scrape together the assured compellence level, which resolves the crisis with  $S_1$ 's capitulation, and the probability of war drops back to zero. In other words, although these types do spend more during the crisis, they obtain the surrender of their opponent without risking war. The possibility to compel  $S_1$  arises out of the latter's initial decision: he could have mobilized enough resources to make himself uncompellable by even the highest valuation type but, because of uncertainty, it is not optimal to do so. This is not to say that technology, war costs, and capabilities are not important—indeed, the two examples show the impact of  $S_2$ 's war costs—but rather that the commonly accepted crisis dynamics based on incentive-compatibility arguments dependent on a fixed distribution of power may not hold when that distribution of power is endogenous.

Furthermore,  $S_1$ 's optimal mobilization is not monotonically related to either his fighting costs or those of his opponent. For example, recall that, when  $c_2 = 0.01$ ,  $m_1^* = 0.25$  in the risk of war equilibrium in Figure 1(b). Increasing  $S_2$ 's costs to  $c_2 = 0.25$  produces  $m_1^* = 0.50$  in an assured compellence equilibrium with no bluffers (figure not shown). Increasing them further to  $c_2 = 0.35$  produces  $m_1^* = 0.07$  in the compellence equilibrium with bluffers in Figure 1(a). Note the distinction between the last two outcomes. When  $S_2$ 's costs are intermediate,  $S_1$  eliminates all bluffers and practically ensures that he would obtain  $S_2$ 's capitulation (the probability of him having to concede instead is less than 1%). When  $S_2$ 's costs increase further,  $S_1$  responds by drastically slashing his own military spending, even exposing himself to bluffing by doing so. Although he is now quite likely to concede (67%), his loss in this case is not too drastic because of the savings from the low allocation. In the previous case, on the other hand, even though he was nearly certain to win, the cost of doing so was quite high, making this tactic no longer profitable. In expectation,  $S_1$ 's payoff does increase in  $c_2$ , and he obtains 0.13 in the latter case as opposed to 0.11 for the intermediate costs case. Perhaps counterintuitively, *the status quo power is more likely to concede when his opponent is weaker (has higher costs of fighting) but equilibrium mobilization levels will be lower.*

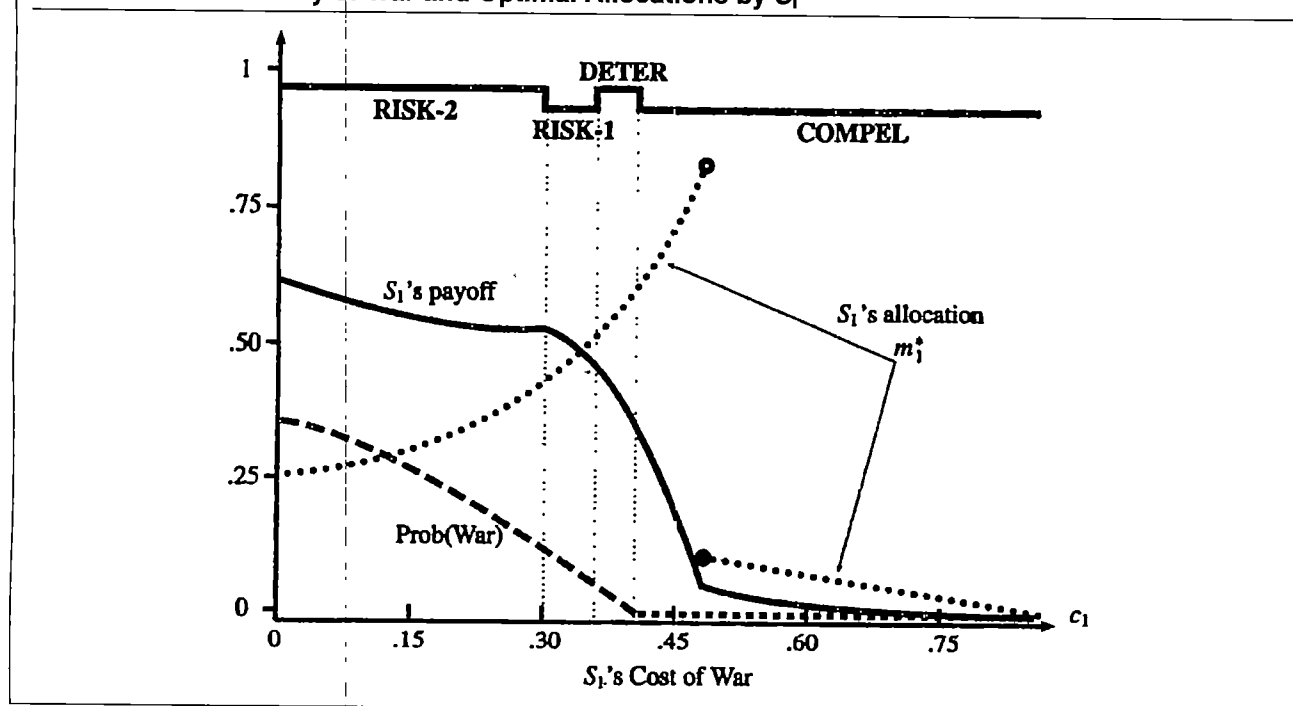
## The Price of Peace

Figure 2 illustrates the impact of varying  $S_1$ 's costs. It shows the ex ante probability of war,  $S_1$ 's optimal allocation, and his payoff in equilibrium for various values of  $c_1$ . The parameters are set to  $v_1 = .999$  (so that high costs do not become immediately prohibitive),  $\lambda = .99$ , and  $c_2 = 0.10$ .

The nonmonotonicity is again evident. Because of his extremely high valuation,  $S_1$  cannot be compelled if his

<sup>14</sup> I thank an anonymous referee for suggesting how to frame this point.

<sup>15</sup> It is true that there exist attributes that make an actor stronger ex ante: large industrial capacity, significant resource stockpiles, sizeable standing army, advanced technology, high-quality training of troops, and mobilization efficiency; all these are enabling characteristics that help an actor gear up for war and can distinguish him from opponents who lack such war-making potential. To this extent, the conventional determinants of strength are fine. However, because most wars are quite short (median duration is less than six months), short-term capabilities matter more than long-term mobilization potential (Huth 1988). Signorino and Tarar (1999) find that the immediate balance of forces has a greater effect than even the short-term balance. Military mobilization of existing forces can affect the expected payoff from war dramatically even if there is a significant resource asymmetry that would render the outcome of a protracted war fairly predictable.

FIGURE 2. Probability of War and Optimal Allocations by  $S_1$ 

costs are relatively low. It is only at intermediate costs ( $c_1 > 0.30$ ) that compellence becomes feasible again. However,  $S_2$  will not attempt it in equilibrium, and hence, up to  $c_1 \approx 0.35$ , war is certain if  $S_2$  mobilizes. The ex ante probability of war declines across this range but  $m_1^*$  increases. That is, seemingly aggressive mobilization behavior can be seen as  $S_1$  compensating for the relative weakness in war occasioned by somewhat high costs: because war is more painful, he is prepared to pay more to decrease the chances of having to fight it. Nothing, of course, can help  $S_1$  overall in the sense that the costlier the fighting, the less must he accept in expectation.

Continuing the increase of  $c_1$  makes assured compellence not just feasible but also desirable, and from  $c_1 \approx 0.35$  no equilibrium outcome will involve war because  $S_1$ 's high costs make fighting quite unattractive for him. Peace can be had in two ways: either  $S_1$  can deter his opponent, or  $S_2$  can compel her opponent.  $S_1$ 's behavior in the intermediate cost range is rather intriguing. While he can afford it, his strategy is to deter  $S_2$  or, failing that, to ensure that the probability of a challenge (to which he will surely concede) is relatively low. Note that, until  $c_1 \approx 0.45$ , the outcome is either assured deterrence or assured compellence but with extremely high mobilization levels by  $S_1$ . Even after it becomes impossible to deter all types of  $S_2$ , the status quo power persists in very high allocations that minimize the probability of having to concede in the compellence equilibrium (less than 0.1%). This is where peace can be very expensive.

Finally  $c_1$  becomes prohibitively high, and  $S_1$  drastically revises his strategy: maintaining a low probability of concession becomes too expensive. The trade-off between the costs of mobilization and expected con-

cessions kicks in, and  $S_1$  precipitously decreases his allocation, exposing himself to ever increasing possibilities for bluffing as his costs go up.

As Figure 1(b) made clear,  $S_2$  types with high valuations must spend substantially more to compel  $S_1$  to capitulate than to fight him. This is, perhaps, not very surprising: given the initial mobilization by the status quo power, it may take a lot of threatening to persuade him to relinquish the prize peacefully. Still, it does go to show that peace can be expensive. This conclusion receives very strong support once we investigate the initial decision itself, as we did earlier. *Peace may involve mobilizations at levels that are substantially higher than mobilizations that precede the outbreak of war. The price of peace can be rather steep either for the status quo state or for the potential revisionist.*

As war becomes costlier,  $S_1$  minimizes the probability of having to wage it, even when this requires skyrocketing mobilization costs. The goal of avoiding war transforms into the goal of avoiding concessions, and  $S_1$  spends his way into successful deterrence until that, too, becomes too expensive. When this occurs,  $S_1$  simply "gives up" and switches to having a permanent, but small, military establishment. That is, he mobilizes limited forces he does not expect to use, and whose impact on the potential revisionist's behavior is rather minimal. These "useless" mobilization levels do serve to weed out frivolous challenges but generally do not work as a deterrent to genuine revisionists or to more determined bluffers.

Peace need not be expensive if either actor has very high costs of fighting. Its price rises steeply, however, when these costs go down. Powell (1993) finds that the peaceful equilibrium in a dynamic model where states redistribute resources away from consumption toward

military uses also involves nonzero allocations, which sometimes can be quite substantial. The results here underscore his conclusions and provide a nuance to their substantive interpretation and empirical implications. These findings further imply that *the common assumption of a costless status quo outcome in formal models may be quite distorting because it fails to account for the resources states must spend on mutual deterrence to maintain it.*<sup>16</sup>

It is worth emphasizing that peace does not depend solely on the credibility of threats. In fact, when war occurs in equilibrium, both actors possess perfectly credible threats and both know it. However, their prior actions have created an environment where neither finds war sufficiently unpleasant compared to capitulation. This illustrates the danger of committing oneself without ensuring that the opponent is not similarly committed (Schelling 1966). Although this may happen easily when actors move simultaneously, it is perhaps surprising that it can also happen when they react sequentially and seemingly have plenty of opportunity to avoid it.<sup>17</sup>

There may exist circumstances where, although peace is, in principle, obtainable, the cost of guaranteeing it is so high that the actors are unwilling to pay it. Peace in this model requires the successful compellence of  $S_1$  or deterrence of  $S_2$ . In a situation where the value of war is determined endogenously, each actor can potentially be coerced into capitulation. The interesting question becomes why sometimes one or both of them choose not to do it. There are, of course, the trivial cases where the cost of doing that exceeds one's valuation so that it is not worth it (assured deterrence), but, more intriguingly, there are the cases where the necessary allocation costs less than one's valuation. In the second example, all types  $v_2 \in (\alpha, \beta)$  fight optimally even though allocating  $\bar{m}_2 = \alpha$  would ensure  $S_1$ 's capitulation.

### Creating Commitments and Communicating Them Credibly

Consider the notion of credibility in the common rational deterrence models.<sup>18</sup> These models postulate a preference between capitulation and war: a resolved actor prefers to fight (and therefore has a credible threat), and an unresolved actor prefers to concede. Some commitments, like an American promise to defend California, are inherently believable, but most are

not (Schelling 1966, 35). This literature has focused on problems with communicating intentions when commitments are not inherently credible. The typical analyses assume that at least one actor is uncertain if its opponent has a credible threat and then investigate how existing commitments can be credibly revealed.

Although superficially analogous, these models are very different from the one presented here because they assume that actors are unable to change the bargaining situation: one either has a credible threat or does not. However, in addition to their informational role, strategic moves can have a functional one (O'Neill 1991). They may alter the physical environment and restructure incentives altogether. That is, bargaining can *create commitments* because actors can manipulate their expected payoffs from following through on a threat and failing to do so.

Consider a stylized scenario where an actor makes a demand and issues a threat to go to war if the opponent does not concede. If that actor restructures the situation such that fighting becomes more attractive than ending the crisis without obtaining the concession, then it has effectively created a commitment not to back down. Imagine scales with the expected payoff from backing down (peace) on one side and the expected payoff from fighting (war) on the other. As long as the peace payoff outweighs the war payoff, the actor has no credible threat. Subtracting weight from the peace side (by making public statements that engage the national honor) or adding weight to the war side (by mobilizing troops) alters the balance, and eventually the war payoff may outweigh the peace payoff: at this point, the actor has created a credible commitment to fight.

Fearon (1994) offers a commitment model of this type. In it, leaders who choose to continue the crisis incur ever increasing *audience costs*; that is, the longer they escalate, the costlier it is for them to back down. If they prolong the crisis sufficiently, they will become locked into positions from which neither would recede, and the inevitable outcome will be war. The basic mechanism that enables them to tie their hands relies on progressively decreasing the benefit of peace until at some point war becomes the more attractive option.

Despite its popularity, the audience cost mechanism has several shortcomings. First, we have had limited success accounting for its microfoundations; that is, the domestic politics that would generate these costs (Schultz 2001b; Slantchev n.d.; Smith 1998). Second, audience costs are not inherently costly because leaders only pay them if they back down without obtaining concessions from their opponents. As Fearon (1997, 80) notes, leaders can generate arbitrarily high audience costs if they want because there are no physical constraints on doing so. Third, the mechanism requires the demanding assumption that leaders incur sufficiently high audience costs; so high, in fact, that peace becomes worse than war. When one considers something as vague and as amorphous as "national honor" and compares it to the destruction of lives and property, and the psychological scars a war inflicts on participants, this assumption becomes heroic indeed.

<sup>16</sup> As a reviewer points out, when models normalize the status quo value to zero, they do not assume that it is costless but that the costs are sunk in the history preceding the game. My model endogenizes military investment and, by showing the effect of its strategic uses, implicitly argues that postulating fixed payoffs for the status quo may be distorting.

<sup>17</sup> Consider the game of Chicken and suppose each player could precommit to standing firm. If precommitment choices are simultaneous, then they may easily end up in a situation where they both precommit to stand firm, making disaster certain.

<sup>18</sup> See Zagare and Kilgour (2000) for an authoritative treatment, with the references therein. Almost all existing models and most of the informal work shares the shortcoming I identify in this section.

Military moves are a suitable candidate for coercive bargaining behavior that has both informational and functional aspects, and they do not suffer from the empirical implausibility of other commitment tactics. To gain some intuition about the workings of the military instrument, consider the other side of the decision-for-peace equation—the expected payoff from war—and actions such as mobilizing troops and sending them to the likely war zone. These are costly activities but they do improve one's chances should war actually break out. Imagine the precrisis situation of insufficient fighting preparedness with the attending prospect of having to spend the resources to “get there.” Compare this with the situation in which one has already paid the costs, and one's troops are ready to go on a short notice. Clearly, the latter situation would afford one a better bargaining position because one's expected payoff from war is now so much higher. If one succeeds in improving that expectation sufficiently, war can become more attractive than peace under the new circumstances, thereby enabling one to commit credibly to fighting. One's military moves can create a credible commitment. Unfortunately, the process of creating and communicating such a commitment may lead to war.

To see how this logic operates, let's examine the example in Figure 1(b) with complete information. Suppose  $v_2 = 0.5$ ; that is, she is one of the types that would end up in a war under incomplete information. It is easy to verify that in the unique subgame perfect equilibrium war does not occur. Instead,  $S_1$  allocates  $m_1^* \approx 0.37$ , and  $S_2$  capitulates immediately. The outcome is successful deterrence by  $S_1$ . What is especially striking about this result is that  $S_1$  achieves deterrence even though his best war-fighting payoff ( $-0.02$ ) is worse than immediate capitulation ( $0$ ). In other words, in a regular deterrence model, this actor does not possess a credible threat, and so one should not expect it to prevail under complete information. Why does this work here? Because sinking the mobilization cost makes capitulation costlier than before: if  $S_2$  resists, the new choice  $S_1$  has is between quitting (which now yields a payoff of  $-0.37$ , the sunk cost of mobilization) and fighting. The payoff from fighting at  $m_1 = 0.37$ , assuming  $S_2$  mobilizes at her optimal level  $m_2^*(0.37)$ , would be at least  $-0.05$ . Thus,  $S_1$  has tied his hands by sinking the mobilization costs at the outset, and he will certainly fight if challenged now even though at the outset he would have capitulated rather than fought even under the best circumstances. Because of  $S_1$ 's rather high mobilization level, fighting becomes too painful for  $S_2$  and so she capitulates. In this way, the military instrument has enabled  $S_1$  to create a credible commitment, and, because there is no uncertainty, the crisis is resolved in his favor.

Under complete information, communicating a commitment is not an issue. Consider now the analogous situation under asymmetric information where  $S_1$  is uncertain about  $S_2$ 's valuation. In this case,  $S_1$  allocates  $m_1^* \approx 0.25$ . First, this is less than what is required to get  $S_2$  with valuation  $v_2 = 0.5$  to capitulate ( $m_1 \approx 0.37$ ). Second, it is more than the maximum mobilization at which

$S_2$  would bother getting  $S_1$  to capitulate ( $m_1 \lesssim 0.23$ ).  $S_1$ 's mobilization level is too high for him to backtrack once  $S_2$ 's valuation is revealed given what  $S_2$  is willing to do, but it is too low to get  $S_2$  to capitulate either. The outcome is war:  $S_1$ 's actions have now created a situation where neither opponent is prepared to back down. This situation arises because of uncertainty and would not have occurred had  $S_1$  known his opponent's valuation from the beginning. Signaling for  $S_2$  is pointless even though it perfectly reveals her valuation, and so her mobilization is simply preparation for war, not a warning.

In the rational deterrence context, the results show that uncertainty drives actors to choose mobilization levels that may change the bargaining context and render capitulation unpalatable to either side despite complete revelation of information. The model demonstrates how this can occur in a two-step fashion: actors fight because they create a situation where they have incentives to do so, and this situation arises because of the actors' crisis behavior under uncertainty. In other words, asymmetric information causes actors to risk committing too much (so that they would not want to back down if resisted) but not quite enough to force their opponent to back down (and so the opponent resists). Military moves may enable one to create and communicate commitments credibly, but, because they are costly and because they can be countered, there are limits to how effective they will be.

The notion of a commitment lock-in under complete information must be tempered: in the model, war occurs without residual uncertainty because the game form does not allow actors to bargain. Hence, the model does not speak to the inefficiency puzzle with complete information (Fearon 1995; Powell 2004). Rather, it provides a rationale for taking the military instrument seriously. Incorporating it in a flexible bargaining context must remain an avenue for future work.

## CONCLUSION

Verbal threats to use force are neither inherently costly nor do they improve one's chances of victory should war break out. In militarized bargaining, threats are implicit in the crisis behavior where actual costs are incurred in activities that could contribute to the success of the military campaign should one come. Hence, military actions can sink costs and tie hands at the same time. I argued that most existing theories of crisis bargaining neglect this dual effect, and consequently their conclusions need to be modified—some substantially, others more subtly. Many empirical hypotheses can be drawn from the preceding analysis. In lieu of enumerating these again, I offer one interesting implication of the overall results.

Fearon (1994, 71) argues that “a unitary rational actor question (how can states credibly signal their foreign policy intentions despite incentives to misrepresent?) proves to require an answer with a nonunitary conception of the state.” This claim is correct if one assumes that military measures involve only sunk costs. However, such an assumption is difficult to sustain on

empirical grounds, and I have shown that, once it is relaxed, unitary actors do recover their signaling abilities. Therefore, there is no a priori reason to privilege domestic politics to explain crisis bargaining.

If actors can use the military instrument to establish credible commitments, and if they are capable of signaling foreign policy through military means, the relative importance of audience costs and other domestic politics mechanisms becomes an open question. In particular, even if such mechanisms operate differently across regime types, there is no reason to expect that they would translate into crisis behavior that would itself depend on regime type. For example, even if democracies are able to generate higher audience costs than autocracies (Fearon 1994), or even if domestic political contestation enables them to reveal more information than autocracies (Schultz 2001a), it does not necessarily follow that democracies would be able to signal their resolve any better in a crisis in which military means are available to autocracies as well. One immediate consequence is that, unless they specify why autocracies forego these signaling possibilities, theories that explain the democratic peace on signaling grounds face a serious difficulty.

Of course, the model also demonstrates that mobilization serves as an implicit threat, and its role as a purely signaling device to warn the opponent of the dangers of escalation may be limited. This suggests that a potentially fruitful theoretical investigation would be to consider the choice between resorting to military moves and sticking to public commitments. Military coercion can be exceptionally dangerous because it alters the strategic environment and may change it to such an extent that war becomes a necessity. Empirically, then, it may not be clear whether mobilization is a warning or a preparatory step to war, a fact that helps explain why it is regarded nervously by crisis participants.

## APPENDIX: PROOFS

**Proof of Lemma 1.** It suffices to show that the maximum expected payoff from fighting is increasing in  $S_2$ 's type at a slower rate than the payoff from assured compellence:  $(\partial W_2^*(m_1, m_2^*(m_1, v_2)))/\partial v_2 = 1 - \sqrt{m_1/\lambda} v_2 < 1 = (\partial [v_2 - \bar{m}_2(m_1)])/\partial v_2$ . Since  $\beta(m_1) - \bar{m}_2(m_1) = W_2^*(m_1, m_2^*(\beta(m_1)))$ , these derivatives imply that  $v_2 - \bar{m}_2(m_1) > W_2^*(m_1, m_2^*(m_1, v_2))$  for all  $v_2 > \beta(m_1)$ . ■

**Proof of Lemma 2.** Suppose  $\delta \geq \alpha$ . The payoff from assured compellence equals zero for type  $\alpha$  while the payoff from optimal war equals zero for type  $\delta$ . Since the expected payoff from assured compellence is strictly increasing in type,  $\delta > \alpha$  must strictly prefer compellence to war. By Lemma 1, it follows that all types  $v_2 \geq \alpha$  strictly prefer assured compellence to both optimal war and capitulation. Hence, if  $\alpha \leq \delta$ , then all  $v_2 < \alpha$  capitulate in equilibrium, and all  $v_2 \geq \alpha$  mobilize at the compellence level. ■

**Proof of Lemma 3.** Suppose  $\delta < \alpha$ . There are three possibilities, depending on where  $\beta$  is located. Suppose  $\delta < \beta < \alpha$ . This implies that all types  $v_2 \geq \beta > \delta$  prefer compellence to optimal war, and war to capitulation, which implies they must prefer compellence to capitulation. But  $v_2 < \alpha$  implies that capitulation is preferred to compellence, a contradiction for

all types  $v_2 \in [\beta, \alpha]$ . Suppose  $\beta < \delta < \alpha$ . This implies that all types  $v_2 \geq \delta > \beta$  prefer compellence to war and war to capitulation, and so they must prefer compellence to capitulation. However, all types  $v_2 \in [\delta, \alpha]$  prefer capitulation to compellence, a contradiction. Suppose  $\delta < \alpha < \beta$ . This is the only possibility that is consistent with the preferences signified by these cut-points. All  $v_2 < \delta$  prefer capitulation to both compellence and war, all  $v_2 \in [\delta, \beta]$  prefer war to both compellence and capitulation, and all  $v_2 > \beta$  prefer compellence to both war and capitulation. ■

**Proof of Proposition 1.** The on- and off-the-path beliefs can be specified as follows: if any  $m_2 < \bar{m}_2$  is observed, update to believe that  $v_2$  is distributed by  $F$  on  $[0, \bar{m}_2]$ , and if any  $m_2 \geq \bar{m}_2$  is observed, update to believe that  $v_2$  is distributed by  $F$  on  $[\bar{m}_2, 1]$ . With these beliefs, if some type  $v_2 < \alpha$  deviates and allocates  $0 < m_2 < \bar{m}_2$ , then  $S_1$  responds by resisting. Since  $\delta \geq \alpha$ , war is worse than capitulation for this type, and so she would capitulate and get  $-m_2 < 0$ , so that such a deviation is not profitable. Allocating  $m_2 \geq \bar{m}_2$  and ensuring capitulation by  $S_1$  is not profitable for this type by construction. Suppose that some type  $v_2 \geq \alpha$  deviated to  $m_2 < \bar{m}_2$ , to which  $S_1$  responds by resisting. Since  $\delta \geq \alpha$ , Lemma 2 implies that such war would be worse than assured compellence. Finally, by the argument in the text, deviation to  $m_2 > \bar{m}_2$  cannot be profitable for any type. Uniqueness follows from Lemma 2, which pins down  $S_2$ 's optimal behavior. It is possible to find other beliefs that would sustain this equilibrium, but they all result in the same behavior. ■

**Proof of Proposition 2.** First, we need to decide what  $S_1$  will believe following an equilibrium mobilization by a nonempty set of  $S_2$  types that has measure zero—that is, when some types mobilize at the same level but the set itself has an equilibrium probability of zero. I assume that the support of  $S_1$ 's beliefs conditional on such mobilization is restricted to the set of types that mobilized at this level. This is necessary because each  $S_2$  type who expects to fight mobilizes at a unique level that is optimal only for that type. What is  $S_1$  supposed to believe after observing such a mobilization? Since there are no atoms in the distribution of types, the probability of any particular type is zero, and Bayes rule does not yield an answer. The restriction requires  $S_1$  to infer the type for whom the given allocation level would have been optimal for war even though only one type would make it in equilibrium.

Assume  $\delta \leq \alpha$  and  $\delta < 1$ . The three cases to consider are  $\alpha < \beta < 1$ ,  $\alpha < 1 < \beta$ , and  $1 \leq \alpha$ . On the path, beliefs are updated via Bayes rule. In particular, for any allocation  $m_2 \in [m_2^*(m_1, \delta), \bar{m}_2]$ ,  $S_1$  infers  $S_2$ 's type with certainty. The off-the-path beliefs can be specified as follows: if any  $m_2 < m_2^*(m_1, \delta)$  is observed, update to believe that  $v_2$  is distributed by  $F$  on  $[0, \delta]$ , and if any  $m_2 \geq \bar{m}_2$  is observed, update to believe that  $v_2$  is distributed by  $F$  on  $[\beta, 1]$  or, if  $\beta > 1$ , any beliefs would work. This equilibrium is unique up to a specification of off-the-path beliefs. ■

**Proof of Proposition 3.** All information sets are off-the-path but any beliefs that  $S_1$  might hold would sustain this equilibrium. Since  $\alpha \geq 1$ , no  $m_2 \leq 1$  can induce  $S_1$  to quit even if he is sure war would occur. Hence, he would resist all such allocations. If any type deviates to such  $m_2$ , war is certain, but  $\delta \geq 1$  implies that even optimal war is worse than capitulation for all types. If any type deviates to some  $m_2 \geq \bar{m}_2 > 1$ , then  $S_1$  would quit for sure but the payoff is strictly negative for all types, and hence such deviation is not optimal. ■

## REFERENCES

- Austen-Smith, David, and Jeffrey S. Banks. 2000. "Cheap Talk and Burned Money." *Journal of Economic Theory* 91 (1): 1–16.
- Banks, Jeffrey S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34 (August): 599–614.
- Banks, Jeffrey S., and Joel Sobel. 1987. "Equilibrium Selection in Signaling Games." *Econometrica* 55 (May): 647–61.
- Bruto, Dagobert L., and Michael D. Intriligator. 1985. "Conflict, War, and Redistribution." *American Political Science Review* 79 (December): 943–57.
- Brodie, Bernard. 1959. *Strategy in the Missile Age*. Princeton: Princeton University Press.
- Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102 (May): 179–221.
- Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (November): 1431–51.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (September): 577–92.
- Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (Summer): 379–414.
- Fearon, James D. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41 (February): 68–90.
- Fudenberg, Drew, and Jean Tirole. 1991. "Perfect Bayesian and Sequential Equilibrium." *Journal of Economic Theory* 53 (2): 236–60.
- Grossman, Sanford J., and Motty Perry. 1986. "Perfect Sequential Equilibrium." *Journal of Economic Theory* 39: 97–119.
- Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (April): 175–200.
- Hirshleifer, Jack. 1988. "The Analytics of Continuing Conflict." *Synthese* 76 (August): 201–33.
- Huth, Paul K. 1988. *Extended Deterrence and the Prevention of War*. New Haven: Yale University Press.
- Iklé, Fred Charles. 1964. *How Nations Negotiate*. New York: Harper & Row.
- Jervis, Robert. 1970. *The Logic of Images in International Relations*. New York: Columbia University Press.
- Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30 (January): 167–214.
- Kydd, Andrew. 2000. "Arms Races and Arms Control: Modeling the Hawk Perspective." *American Journal of Political Science* 44 (April): 222–38.
- Morrow, James D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33 (November): 941–72.
- Morrow, James D. 1994a. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38 (2): 270–97.
- Morrow, James D. 1994b. "Modeling the Forms of International Cooperation. Distribution versus Information." *International Organization* 48 (Summer): 387–423.
- O'Neill, Barry. 1991. "Conflictual Moves in Bargaining: Warnings, Threats, Escalations, and Ultimatums." In *Negotiation Analysis*, ed. H. Peyton Young. Ann Arbor: The University of Michigan Press. Chapter 5.
- Powell, Robert. 1990. *Nuclear Deterrence Theory: The Search for Credibility*. Cambridge: Cambridge University Press.
- Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87 (March): 115–32.
- Powell, Robert. 1999. *In the Shadow of Power*. Princeton: Princeton University Press.
- Powell, Robert. 2004. "The Inefficient Use of Power: Costly Conflict with Complete Information." *American Political Science Review* 98 (May): 231–41.
- Quester, George. 1977. *Offense and Defense in the International System*. New York: Wiley.
- Rector, Chad. 2003. "Getting There is Half the Battle: How Military Mobilization Costs Influence Crisis Bargaining." Presented at the 2003 Meeting of the Midwest Political Science Association. Chicago IL.
- Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (February): 121–49.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven: Yale University Press.
- Schultz, Kenneth A. 2001a. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.
- Schultz, Kenneth A. 2001b. "Looking for Audience Costs." *The Journal of Conflict Resolution* 45 (February): 32–60.
- Signorino, Curtis S., and Ahmer Tarar. 1999. "A Unified Theory and Test of Extended Immediate Deterrence." Presented at the Annual Meeting of the American Political Science Association, Washington, DC.
- Slantchev, Branislav L. 2003. "The Principle of Convergence in Wartime Negotiations." *American Political Science Review* 47 (December): 621–32.
- Slantchev, Branislav L. 2004a. "Mobilization, Offense Dominance, and Preemptive War." University of California, San Diego Manuscript.
- Slantchev, Branislav L. N.d. "Politicians, the Media, and Domestic Audience Costs." *International Studies Quarterly*. Forthcoming.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
- Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87 (August): 355–74.
- Strachan, Hew. 2003. *The First World War*. New York: Viking.
- Zagare, Frank C., and D. Marc Kilgour. 2000. *Perfect Deterrence*. Cambridge: Cambridge University Press.



# Contracting around International Uncertainty

BARBARA KOREMENOS *University of Michigan*

*International cooperation is plagued by uncertainty. Although states negotiate the best agreements possible using available information, unpredictable things happen after agreements are signed that are beyond states' control. States may not even commit themselves to an agreement if they anticipate that circumstances will alter their expected benefits. Duration provisions can insure states in this context. Specifically, the use of finite duration depends positively on the degree of uncertainty and states' relative risk aversion and negatively on the cost. These formally derived hypotheses strongly survive a test with data on a random sample of agreements across all four of the major issue areas in international relations. Not only do the results, highlighting evidence on multiple kinds of flexibility provisions, strongly suggest that the design of international agreements is systematic and sophisticated; but also they call attention to common ground among various subfields of political science and law.*

In an international environment characterized by uncertainty that is at times so pronounced that it could easily make states fearful of making meaningful commitments, how is cooperation possible? States often have large amounts of information at their disposal when initiating cooperative activity and use it wisely to set the terms of cooperation and to manage its evolution. Inevitably, though, events occur that could not have been foreseen and cannot be controlled. Why do states commit themselves to cooperation in the first place if they expect that circumstances might reduce their anticipated benefits? Not only do states cooperate; they also codify their cooperation through countless international agreements that form a substantial body of international law. How is this possible?

One answer is that perhaps cooperative efforts are undertaken only in the areas and among states for which uncertainty is less pronounced. This may very well be the case given that, according to the conventional wisdom in international relations, a main objective of international cooperation is to make states' commitments to each other more credible through hands-tying. If we acknowledge both the conventional wisdom and the pervasive uncertainty in the international environment, we might conclude that, because states do not like to tie their hands under conditions of high uncertainty, cooperation ensues only when uncertainty is low.

Does this imply that countless other cooperative possibilities go unrealized because states cannot "in-

sure" themselves against the unanticipated negative consequences that uncertainty might bring? Or should we entertain the possibility that states can somehow protect themselves in situations of high uncertainty?

At first glance, this second possibility seems far-fetched. After all, international cooperation takes place in anarchy, so the idea that international cooperation could be designed to provide some kind of insurance for states is hard to imagine. On the other hand, the need for such a scheme is especially pressing in international relations given the pronounced uncertainty that operates in that sphere. I show that such a scheme is possible, notwithstanding the complications imposed by anarchy.

Taking as my point of departure economic contracting theory, I develop a model that features flexibility rather than hands-tying to make such protection possible. The specific flexibility provision that serves as an "international insurance scheme" is a limited duration agreement that can be renegotiated.<sup>1</sup> I subject the model to empirical testing with data on a random sample of agreements drawn from the United Nations Treaty Series (UNTS). This is the first study that uses a random sample across all four of the major issue areas in international relations to examine matters of international agreement design.<sup>2</sup>

This study also forces us to consider whether international relations scholars are missing opportunities by not examining the extent to which there is common ground with scholars in some of the other subfields of political science. More specifically, unlike in international relations, in much of American and comparative politics the study of institutions and institutional design has gained prominence, with much of that work drawing on economic theory. If international law follows a logic similar to one developed for economic contracts, might there not be much more in common among the various subfields of political science than traditionally

Barbara Koremenos is Associate Professor of Political Science, University of Michigan, 5700 Haven Hall, 505 South State St., Ann Arbor, MI 48109-1045 (koremeno@polisci.ucla.edu).

This article draws on research carried out under a National Science Foundation CAREER Award: "Designing International Agreements: Theoretical Development, Data Collection, and Empirical Analysis" (SES-0094376). Several colleagues gave me helpful comments on various drafts of this work. In particular, I thank Beth Simmons, Matt Baum, Jerry Cohen, Ron Mitchell, Rip Smith, Jim Morrow, Bob Powell, David Lake, Duncan Snidal, and Paul Diehl. I thank seminar participants at Chicago, Columbia, and UCSD for their feedback, and Jana von Stein for valuable research assistance. The comments of editor, Lee Sigelman, and the three anonymous referees greatly improved this article. Finally, I thank Jeffrey Smith for many stimulating discussions as well as detailed comments on numerous drafts of this paper and George Tsebelis for helping me see the value of my work in a broader context.

<sup>1</sup> In models of the domestic labor context, such flexibility can insure actors against certain kinds of price shocks.

<sup>2</sup> Although this is the first study to use a random sample to examine issues of agreement design, there are a number of analyses featuring careful large-n empirical work in the field of international cooperation, including Simmons 2000, Guzman and Simmons 2002, and Martin 1992, 2000.

acknowledged? Given that over 50 thousand international agreements are ready for us to analyze, the results of this analysis can help guide both future research within international relations and collaborative efforts across subfields and even disciplines.

## EXPLAINING THE STRUCTURE OF INTERNATIONAL AGREEMENTS

This article adds to a small but growing theoretical literature on how states choose the flexibility provisions of international agreements. A few scholars consider an environment in which a state's gain from an agreement is affected by infrequent, transitory negative shocks. In the area of monetary policy under the gold standard, Bordo and Kydland (1995) show how states used "escape clauses" consisting of temporary suspensions of convertibility to deal with banking crises. Downs and Rocke (1995) and Rosendorff and Milner (2001) consider similar escape-clause models in a trade context, where the shocks represent the sudden mobilization of domestic interest groups in response to particular aspects of trade agreements.

A few economists, for example, Gray (1978), Dye (1985), and Harris and Holmstrom (1987), consider an environment in which some underlying random variable affects the payoffs to the parties from a contract under the assumption that the parties cannot make the contract contingent on this variable. These models focus on how the parties choose the optimal duration of their contracts in the presence of uncertainty.

In earlier work, I have considered an environment in which a cooperative endeavor is characterized by a one-time distributional uncertainty and states have to learn about it over time (Koremenos 2001). In that context, states may limit the duration of their agreement and renegotiate it once after enough information has been revealed about the initial uncertainty. My case studies included the Nuclear Non-Proliferation Treaty, for which the relatively fixed distribution of gains between the nuclear weapon states and the non nuclear weapons states was initially unknown. While drawing on the economic models, that analysis adds some refinements (including the ability of states to renege on their contracts) to better capture the essential features of the international environment.

The model presented here builds on my prior analysis, with one key difference: the uncertainty is persistent; hence, there is no learning. Thus, the model here generalizes the earlier results by considering a different kind of distributional uncertainty and, in response, a *series* of renegotiations.

## The Theory

To characterize *persistent* uncertainty surrounding the distribution of future gains from an agreement, I assume that the parties select an initial distribution of gains based on their relative bargaining power, but this gain evolves over time under the agreement due to external shocks. The precise nature of the shocks

depends on the issue area but could include climatic fluctuations (e.g., commodity agreements), exogenous private sector shocks (e.g., exchange rates), or political shocks (e.g., refugee agreements). I assume that the parties always know the distribution of gains in the current period, but know only the probability distribution for the distributions of gains in all future periods. The distribution of gains in each period equals the distribution in the previous period plus a random shock. These random shocks are cumulative, so the distribution of gains, left to its own devices, follows a random walk.

Put simply, a shock to the distribution of gains occurs in each period. As a result, either the gains to one state increase relative to that of the other, or there is no change if the shock equals zero. These shocks have three important features. First, they are completely random. Each shock is independent of every other, both before and after. They are also independent of everything else that the states can observe, which implies that they cannot be predicted, either at the time an initial agreement is made or later on. Second, the precise value of any shock is discoverable only at a cost. In other words, shocks are noisily observable. Third, the shocks add up over time. That is, the gain for a state in any one period is the sum of its initial gain plus each of the subsequent changes. Because the random shocks are cumulative, it is possible for the states' relative shares to evolve in a way that differs quite substantially from the assumed equal division at the start of the agreement. How great the differences become depends both on how much time has passed and on the specific sequence of realized shocks. As time passes, the potential divergence from the initial equal division increases because the passage of time allows, but does not require, the random walk process to walk farther away from its starting point. Hence, once a policy course is chosen, without subsequent redirection, the system may continue down particular paths longer than the parties originally intended. I refer to this environment as one characterized by "persistent uncertainty."

The extent or importance of the *uncertainty* is summarized by the variance of the shocks to the distribution of gains, with a larger variance implying greater uncertainty. Also, the larger the variance, the farther the distribution of gains will depart, on average, from its initial value over a given time.

Of course, states do not embrace such uncertainty; rather, the idea that they may gain far less than they anticipated from an agreement could make some of them too nervous to commit to one. To capture this, I employ a standard assumption in international relations that states are *risk averse*. I do, however, allow them to vary with respect to their *relative* level of risk aversion (see O'Neill 2001).

## Types of Agreements

Suppose states determine that there are gains from cooperation and decide to negotiate an agreement. They have two options in a given agreement context:

(1) an agreement of indefinite duration (an inflexible agreement) or (2) a series of agreements renegotiated at regular intervals in order to adjust the distribution of gains for the effects of the shocks that cumulate during each agreement (a flexible agreement).

It is important to note a few alternatives that are not available. I exclude a completely contingent agreement, consistent with my assumption that it is costly to uncover the exact value of the shock to the distribution of gains. Most international agreements are multidimensional and are characterized by at least some gains that are not easily measurable; hence, it is a complex process to assign values to all possible outcomes. Thus, once a shock has occurred, some kind of meeting among the member states is usually necessary to bring together and analyze all of the information each has individually gathered. Such a meeting enables comparisons across an agreement's multiple dimensions regarding the true winners and losers. Because of the complexity of some of the issue areas, third-party "neutral" expertise is at times solicited. For example, in the environmental issue area, experts are sometimes invited to testify at meetings of the member states.

I also exclude the delegation of flexibility to an international organization, an option found in Koremenos 2000, which models explicitly the creation of an international organization as a choice variable. Current data limitations do not allow for testing the fuller model. In random samples, there is a predominance of bilateral agreements, which rarely call for the creation of international organizations.<sup>3</sup>

Consider first an indefinite duration agreement. States pay initial negotiation costs, but no additional costs thereafter. Given that there is no provision for renegotiating the distribution of gains in response to shocks, the distribution may wander well away from its initial level.

There is no international authority to enforce agreements. Hence, should the distribution of gains move sufficiently far enough away from the initial level, one state may break the agreement. The state that elects to do this pays a cost. The costs of reneging consist of any sanctions imposed by the other parties or by other international actors. These costs may be substantial. I assume that once reneging costs are paid, the states begin again with a new agreement with an indefinite duration. Put differently, I assume that the reneging state incurs no cost in terms of reputation or refusal to bargain, thereby avoiding problems of renegotiation-proofness.

Instead of one indefinite duration agreement, states can choose to conclude a series of finite duration agreements of equal duration, with renegotiation taking place between each pair of agreements in the series. I consider only series of agreements that are of the same duration because of my assumption that the environment facing the parties does not change over time. As a result of this assumption, the states will make the same choice regarding agreement duration at the start

of the first agreement and every subsequent agreement because the choice problem they face in each case does not change.

States choosing to conclude a series of finite duration agreements incur two kinds of cost. First, states incur the negotiation costs required to reach the initial agreement in the series. I assume that these are the same as in the case of a single agreement of indefinite duration, although they could be less if a shorter shadow of the future results in lower bargaining costs. Second, states incur *renegotiation costs* at the start of each new agreement in the series. Like negotiation costs, renegotiation costs vary depending on the states in the agreement and the environment. For example, I expect both costs to increase with the number of parties to the agreement.<sup>4</sup> In addition, with respect to renegotiation costs, the greater number of parties produces greater amounts of information that have to be analyzed to discover both the nature of the shocks and how the terms of the agreement must be adjusted to cancel out their cumulative effect.

The advantage that states derive from concluding a series of renegotiated agreements rather than one agreement of indefinite duration is flexibility: the division of gains can be *reset* to the initial level at regular intervals. States choose to reset the division of gains to its original level because I assume that bargaining power and other factors that influence the division do not change over time.

The planned readjustment to the division of gains that occurs under a series of renegotiated agreements greatly reduces the chance that either state will want to incur the costs of reneging or be forced to endure an unsatisfactory division of gains for long periods. Put differently, for risk-averse states, the opportunity periodically to reset the distribution of gains back to its initial level increases the *ex ante* value of a series of renegotiated agreements relative to a single indefinite agreement. It does so by reducing the variance of the discounted expected gain from the agreement. States that choose to renegotiate must decide how often to do so. Doing so involves trading off the costs of more frequent renegotiation against the costs of living with a distribution of gains that differs from that initially chosen in light of the parties' relative bargaining power.

Foreshadowing the formal results, states will choose to renegotiate more often when *renegotiation costs* fall. States will also renegotiate more often when the *variance of the shocks* increases because this increases the expected deviation between the realized and the desired distributions of gains and when their level of *risk aversion* increases because the variance is more costly to them.

Suppose that the variance of the shocks to the distribution of gains is very low so the benefits from flexibility are small and renegotiation costs are high. In this

<sup>3</sup> The UNTS data at the time my sample was drawn (1999) consisted of 32,939 bilateral agreements and only 2,330 multilateral ones.

<sup>4</sup> Conybeare (1985) makes this argument in relation to trade talks, noting that the multilateral Kennedy Round (1963–67) took considerably longer than the previous bilateral Dillon Round (1960–61).

case, states may choose an indefinite agreement. The 1967 Outer Space Treaty, which forbids the placement of weapons of mass destruction into orbit, onto celestial bodies, or in outer space, is not subject to great uncertainty of this kind. There are no supply-and-demand shocks, and no resource or territorial issues are at stake. Rather, the goal is to lock in the status quo and prevent the militarization of space. The states involved therefore chose not to incur renegotiation costs; the agreement is of indefinite duration.

By contrast, if the variance of shocks to the distribution of gains is large so that it quickly departs from its initial level, states may want to limit the duration of their agreement. The Group of 7 (G-7) macroeconomic cooperation is characterized by low renegotiation costs, given the small number of parties and the low-cost availability of economic information, and by a high variance of shocks to the distribution of gains, given a rapidly changing world economy. Thus the G-7 has traditionally chosen a series of short, renegotiated agreements. In the case of the Group of five finance ministers, costs are lower, the variance of the shocks is higher, and agreements are even shorter.<sup>5</sup>

## Formal Model

I consider two states, State 1 and State 2. Let  $Y_{1,t} = b_{1,t}$ ,  $Y_{2,t} = b_{2,t}$ , denote the outcomes for the two states in the absence of an agreement. The first subscript indicates the state, and the second indicates the time period in each case. These outcomes depend on the specific issue area under consideration, but may include things like GDP, trade levels, or exchange rate stability, appropriately measured. For simplicity, I assume that the base outcomes do not change over time.

In the presence of an agreement, the outcomes attained by the two parties to an agreement in each period  $t$  are given by  $Y_{1,t} = b_{1,t} + m_t$ ,  $Y_{2,t} = b_{2,t} + (g - m_t)$ , where  $g$  is the total gain from the agreement (the size of the pie) and  $m_t$  is the portion received by State 1. For simplicity, I assume a fixed total gain  $g$ . Rather than having the total gain vary over time, I have only the division of gains between the parties vary over time. Although this sacrifices some realism, it simplifies the model and allows me to focus on the key issue of how states structure agreements in the face of distributional uncertainty.

To take account of the rapidly changing context of international agreements, I assume that in the absence of renegotiation, the distribution of gains evolves according to a random walk, with  $m_t = m_{t-1} + e_t$ , where  $m_0$  is chosen under the agreement and where  $e_t$  has density function  $h(e_t)$ . I assume, without loss of generality, that

the states are symmetric. This implies equal bargaining power and, under a Nash bargaining solution, equal initial shares, so that  $m_0 = 0.5g$ .<sup>6</sup>

I assume that  $E(e_t) = 0$  and that  $e_t$  is independently and identically distributed across periods. The parties observe  $m_t$  in each period. Let  $t = 0$  denote the first period of an agreement. By choosing  $m_0$ , the parties to the agreement are choosing the expected value, as of  $t = 0$ , of  $m_t$  for all future periods. Because the  $e_t$  are independent across periods and have a common variance, we have

$$m_t = m_0 + e_1 + e_2 + \dots + e_t, \quad \text{and} \\ \text{var}(m_t) = \text{var}(g - m_t) = t * \text{var}(e_t).$$

That is, because the shocks are independent and cumulative, the set of possible distributions of gains from the agreement fans out over time as the agreement continues. Put differently, the probability that the realized value of the distribution of gains differs by any fixed amount from the initial choice of  $m_0$  increases over time. In the general case considered here, the gain for one party (either  $m$  or  $(g - m)$ ) from the agreement may become negative (though an agreement whose initial gain was negative for one or both parties would, obviously, not elicit much interest).

I denote the cost of negotiating a single, infinite duration agreement or the first in a series of renegotiated agreements by  $k_n$ . When an agreement gets renegotiated, states pay a renegotiation cost  $k_r$ . These costs are incurred by all of the parties to an agreement. I interpret both  $k_n$  and  $k_r$  much more broadly than is done in the economics literature. In addition to the costs of sitting down and talking, I treat both as including the costs of assembling or reassembling domestic political coalitions—often necessary for agreement ratification. Renegotiation costs also include the costs of discovering the value of  $e_t$ .

As noted earlier, when states renegotiate an agreement, they reset the distribution of gains under the agreement to that originally agreed upon. This follows from my assumption of stable relative bargaining power. This means that  $m_t = m_0$  in periods immediately following a renegotiation.

To focus on the basic tradeoffs, consider a simple two-period case. The parties must choose between one 2-period agreement (an inflexible agreement) and two 1-period agreements (a flexible agreement in that it can be renegotiated). If they choose two 1-period agreements, each gains a reduction in the variance of its gain from  $2 * \text{var}(e_t)$  to  $\text{var}(e_t)$ , but loses the additional renegotiation cost. The expected value (gross of negotiation and renegotiation costs) is the same in both cases; consequently risk-neutral parties would opt for the 2-period agreement, as variance does not matter to them.

<sup>5</sup> It could be argued the more an agreement matters, the more important is flexibility. In fact, why would states pay for flexibility in a shallow or trivial agreement? An additional variable measuring importance should perhaps be interacted with the uncertainty variable. Unfortunately, that measure does not exist and would be extremely difficult to create in any objective sense.

<sup>6</sup> This cooperative solution corresponds to the Rubinstein alternating-offers noncooperative solution when  $\delta$  is close to 1. See Osborne and Rubenstein 1990.

Formally, the parties compare their expected utilities with one 2-period agreement, given by

$$E(u(b_1 + m_0 + e_1 - k_n) + \delta u(b_1 + m_0 + e_1 + e_2)), \\ E(u(b_2 + (g - (m_0 + e_1)) - k_n) \\ + \delta u(b_2 + (g - (m_0 + e_1 + e_2))))),^7$$

to their expected utilities from two 1-period agreements, given by

$$E(u(b_1 + m_0 + e_1 - k_n) + \delta u(b_1 + m_0 + e_2 - k_r)), \\ E(u(b_2 + (g - (m_0 + e_1)) - k_n) \\ + \delta u(b_2 + (g - (m_0 + e_2)) - k_r)).$$

The parties choose the agreement type that provides the higher expected utility.

### Comparative Statics

The following hypotheses implicitly hold everything else equal so they indicate the effect of changing one parameter of the model holding the others constant. The proofs are in Appendix A. These hypotheses focus on the effects of variation in states' characteristics and in the agreement context on the type of agreement selected. I also present hypotheses regarding the choice of agreement duration, should states choose to conclude a series of finite duration agreements. I have also simulated a version of the model that includes the possibility of reneging, using a variety of values for the key parameters: the level of risk aversion, renegotiation costs, and the variance of the shocks to the distribution of gains. The simulations support the comparative statics (see Appendix B).<sup>8</sup>

- (CS-1) As **renegotiation costs** increase, the probability that the parties will choose finite, renegotiated agreements decreases.
- (CS-2) If the parties conclude a series of renegotiated agreements, as **renegotiation costs** increase, parties will choose to make each agreement in the series longer.

An increase in renegotiation costs (CS-1) raises the costs of choosing a series of renegotiated agreements relative to an indefinite duration. With respect to (CS-2), the costs of renegotiation increase while the benefits of renegotiation remain unchanged; thus, the parties will renegotiate less frequently by making each agreement in the series longer.

<sup>7</sup> This formulation implicitly assumes neither party reneges in the two-period case. I assume the possible values of the agreement shock are such that, in only two periods, the distribution of gains never moves far enough away from the initial value to make paying the cost of reneging worthwhile. This is a reasonable assumption except for (hypothetical) states on the extreme margin that completely discount future benefits. Still, the complete case with reneging is described in Appendix B.

<sup>8</sup> These comparative statics also confirm the robustness of the logic articulated in Koremenos 2001 when the uncertainty variable is changed

- (CS-3) As **uncertainty** increases, the probability that the parties will choose finite, renegotiable agreements to adjust for shocks increases.
- (CS-4) If the parties conclude a series of renegotiated agreements, as **uncertainty** increases, the parties will choose to make each agreement in the series shorter.

An increase in uncertainty (CS-3) makes the parties value flexibility more. More specifically, it increases the variation in realized outcomes under an indefinite duration agreement, which makes it less attractive relative to the alternative. With respect to (CS-4), the value to the parties of renegotiating more often to undo the shocks increases; hence, they conclude shorter agreements.

- (CS-5) As the **risk aversion** of the parties increases, the probability that they will choose finite, renegotiable agreements to adjust for shocks increases.
- (CS-6) If the parties conclude a series of renegotiated agreements, as the **risk aversion** of the parties increases, the parties will choose to make each agreement in the series shorter.

In (CS-5), as the parties become more risk averse, they increasingly value some form of flexibility in their agreement to reduce the variation of the realized outcomes. In the case of (CS-6), the intuition is that as the level of risk aversion increases, so do the costs of putting up with an agreement whose distribution of gains has moved far from that originally agreed on. As renegotiation costs are constant, states will choose to renegotiate more often so that the variance in the realized outcomes falls.

Subject to two caveats, the two-period model generalizes readily to an infinite horizon, particularly under the assumption of a time-homogeneous environment. States face the same two choices and again compare their discounted expected utility from each one, choosing the one with the higher value.<sup>9</sup>

The first caveat concerns the alternative of a series of finite duration agreements. In the two-period model, these agreements can only be of length one. In an infinite horizon framework, the states must choose their preferred duration from the set of all possible finite durations by comparing the discounted expected utilities associated with series of finite duration agreements of various lengths. Once they have identified the finite duration agreement length they like best, they can compare it to their discounted expected utility from the other three agreement type choices.

The second caveat concerns the possibility of reneging. Over an infinite time horizon, the distribution of gains under an indefinite agreement could move very far away from the parties' preferred division. If it moved far enough, one party might be losing more

<sup>9</sup> Generalizing the model for an environment that changes over time would be conceptually simple but notationally burdensome and not add anything to the substantive results.

from the agreement than the sum of the expected future gains and the costs of reneging. The same condition could hold in the midst of a relatively long finite duration agreement. Thus, reneging looms larger when the time horizon is extended.

The generalization to additional states is straightforward, subject to the following caveats. First, adding additional states will affect certain parameters of the model. As a result, holding the other parameters constant, agreement type choice, as well as the choice of whether to have an agreement at all, may change with an increase in the number of parties. For example, increasing the number of states will increase renegotiation costs. Moreover, generalizing the model both across time and to multiple parties calls attention to the relationship between potential changes in membership and the anticipated costs of renegotiation. If the agreement is such that membership is likely to increase dramatically, states may not choose renegotiation even though the initial membership is low. Given that membership rules themselves affect the number of future signatories to an agreement, this raises the issue of the interaction between two aspects of institutional design: flexibility and membership rules.<sup>10</sup>

## EMPIRICAL RESULTS

### Data

This article is the first to exploit a new data set, the "Continent of International Law" (COIL).<sup>11</sup> According to Article 102 of the Charter of the United Nations, "every treaty and every international agreement entered into by any Member of the United Nations after the present charter comes into force shall as soon as possible be registered with the Secretariat and published by it" (<http://www.un.org/Overview/Charter/contents.html>). Given the almost universal membership of the United Nations and its stature among international organizations, its list of international agreements is the most comprehensive to be found.<sup>12</sup> All international agreements registered or filed and recorded with the Secretariat since 1946 are published in the United Nations Treaty Series (UNTS). The Internet collection at the time the sample was drawn contained over 34,000 international agreements "which have been published in hard copy in over 1,450 volumes, which corresponds to all treaties and subsequent actions registered up to December 1986" (<http://www.un.org/Depts/Treaty/>).

<sup>10</sup> If I were to allow the bargaining power of the parties to be affected by the realized distribution of gains, this would reduce the expected utility of a renegotiated agreement for all values of the variance of the agreement shock. This would not change any of the comparative static results; rather it would only change the cut point at which states switch from one type of agreement to the other.

<sup>11</sup> COIL is supported by a 5-year National Science Foundation CAREER Award. "Designing International Agreements: Theoretical Development, Data Collection, and Empirical Analysis" (SES-0094376).

<sup>12</sup> The key exception to this completeness is the absence of informal agreements.

The UNTS Internet site provides subject terms that can be used in searching for international agreements. The primary data for this paper are drawn from the monetary matters, finance, investment, agricultural commodities, environment, human rights, security, and disarmament subject headings. Within each issue area, I generated a random sample of the total set of agreements. Once an agreement was selected, it was subjected to screening criteria. I excluded agreements that only established procedures for other agreements or designated the host for international conferences, that did not include at least two states among the parties, and that did not prescribe, proscribe, or authorize behavior that was observable at least in principle. Agreements are also subjected to rules to avoid double counting.<sup>13</sup>

The characteristics of the selected agreements were recorded using a special coding form. Detailed information on flexibility provisions such as duration and renegotiation provisions, withdrawal and escape clauses, and the presence of quasi-legislative institutions designed to adapt to shocks as well as on a number of other design features like membership provisions and voting rules was coded.

### Variables

The dependent variables, whether an agreement is finite or not and if so, how long it is, were measured directly by coding the agreements. Still, it is necessary to operationalize the following independent variables: renegotiation costs, uncertainty, and risk aversion. One reason there has been negligible testing of formal models of international cooperation is the difficulty of finding measures for commonly used variables, like uncertainty and risk attitude. Still, we must rise to the challenge or some of the claims made by critics of the formal approach will be substantiated (Brown 2000). Although I qualify my measures later, they represent earnest attempts to capture the theoretical concepts and thereby serve as useful starting points for much-needed testing in the subfield.

I use the number of original signatories to the agreement as a proxy for *renegotiation costs* because, according to bargaining theory, increasing the number of actors involved is likely to make the negotiation process more lengthy given the existence of multiple equilibria. Also, the amount of information that needs to be analyzed and agreed on before an existing agreement can be adjusted increases with the number of parties involved.

For *uncertainty*, the coding rules are based primarily on categories of agreements in each issue and sub-issue area. In the model, the uncertainty surrounds the variance of shocks to the distributions of gains from the agreement. Certain kinds of agreements are far more subject to these shocks than are others. For example, agreements whose distribution of gains is

<sup>13</sup> See <http://www.polisci.ucla.edu/faculty/koremenos> for the set of agreements in the random sample.

governed by the forces of supply and demand are coded as high uncertainty, whereas those primarily about coordinating policies to avoid suboptimal outcomes are coded as low uncertainty. Put differently, agreements for which changes in the environment can cause the distribution of gains to vary substantially over time, even while being welfare-enhancing in the aggregate, are high-uncertainty agreements, whereas those for which efficiency concerns dominate and period-to-period changes in the distribution of gains are not expected are low-uncertainty agreements. Two scholars weighed in on each decision, often supplemented by conversations with scholars specializing in the issue area, to avoid subjectivity as far as possible.

Beginning with economics, the subissue area of *monetary* agreements essentially contains exchange rate agreements. These are subject to supply-and-demand shocks that could dramatically alter the distribution of gains from period to period. They are all coded as having high uncertainty. All *trade* agreements are coded as high-uncertainty observations for the same reason. *Finance* agreements are of two types: those that resemble monetary or trade agreements—high uncertainty—and those that are about coordinating policies (e.g., Convention for the Avoidance of Double Taxation and the Prevention of Fiscal Evasion with Respect to Taxes on Income between Australia and Italy)—low uncertainty. Almost one third of finance agreements are high uncertainty. *Investment* agreements concern the promotion and protection of investments against nationalization and expropriation and thereby are subject to political shocks that could alter the distribution of gains; hence, they are high-uncertainty agreements.

In the *environmental* issue area, the following rules were followed: agreements addressing plant and bird protection or scientific cooperation on subjects like soil are coded as low-uncertainty agreements (e.g., the Exchange of Notes Constituting an Agreement on the Project Soil Management and Conservation in East Amazonia between Brazil and the Federal Republic of Germany). Such agreements are predominantly about coordinating policies. On the other hand, agreements about pollution abatement or fishing or other sea resources are coded as high uncertainty (e.g., the multilateral International Convention for the Conservation of Atlantic Tunas). Pollution control and sea resources implicate competitive industries. Shocks affecting the availability of or dependence on the resource (fish) as well as technologic shocks (e.g., in pollution control, expected positive developments may not be forthcoming) can alter the distribution of gains. Almost half of the environmental agreements are coded as high uncertainty.

For *human rights*, universal declarations (e.g., the multilateral Convention on the Prevention and Punishment of the Crime of Genocide) are coded as low uncertainty agreements. Such agreements serve to capture or establish ethically based international norms. On the other hand, agreements about refugees and detailed labor standards are subject to distributional shocks. For instance, political shocks may dramatically

change the flow of refugees and thereby change the distribution of gains. Just over half of human rights agreements are coded as high uncertainty.

*Security* agreements fall into two basic categories: universal prohibitions (e.g., the multilateral Agreement Governing the Activities of States on the Moon and Other Celestial Bodies) and those related to mutual security. Following the logic of human rights, prohibitions are low uncertainty, whereas mutual security agreements are high uncertainty. Almost two thirds of security agreements are high uncertainty.

*Risk aversion* is a standard assumption in theoretical models of international relations, but little attention has been paid to developing measures of it and incorporating them into analyses of international cooperation. I use three different proxies, each of which captures some aspect of relative risk aversion.

I turn first to Bueno de Mesquita's (1985) risk attitude measure, a "*de facto* academic standard" (Bennett and Stam 2000a, 541) in the international conflict literature. The measure is based on the premise that the closer a state is to the alliance portfolio that maximizes its security, the more risk-averse it is.<sup>14</sup> To generate risk attitude scores, I used the *EUGene* risk attitude variable (Bennett and Stam 2000b).<sup>15</sup> Risk scores are region-based, ranging from  $-1$  (very risk-averse) to  $+1$  (very risk acceptant). For bilateral treaties, I use regional risk scores. For example, if states  $i$  and  $j$  sign a treaty, I calculate  $i$ 's risk attitude toward state  $j$ 's region and  $j$ 's risk attitude toward  $i$ 's region. For multilateral treaties, which include up to 119 signatories, I compute global risk scores (the mean of each signatory's regional risk scores) for each signatory. The risk attitude assigned to an agreement is a function of the risk attitude of the *most* risk-averse participant.<sup>16</sup> Finally, I convert the risk attitude score into a risk aversion measure by inverting the scale to run from  $-1$  (least risk-averse) to  $+1$  (most risk-averse).

Bueno de Mesquita's (1985) risk attitude variable is not above criticism. One problem is that it assumes that alliances work exclusively to make states more secure and therefore reflect risk aversion; it fails to capture the possibility that alliances increase autonomy

<sup>14</sup> More specifically, the variable is constructed in three steps: (1) define state  $i$ 's "security level" (the sum of all other states' expected utilities vs.  $i$ ); (2) identify the hypothetical alliance portfolio that would minimize and the portfolio that would maximize  $i$ 's security; (3) how proximate  $i$ 's actual policies are to its hypothetical policies may be interpreted as an indication of  $i$ 's willingness to take risks. Bueno de Mesquita (1985, 157) assumes that " $i$ 's risk acceptance increases as  $i$ 's security score approaches its level of greatest vulnerability, and that  $i$ 's risk aversion increases as its security approaches the level possessed by its safest policy portfolio.

<sup>15</sup> According to Bennett and Stam (2000a, 466), *EUGene* uses an improved and more accurate algorithm to generate risk attitude scores than does Bueno de Mesquita (1985). I employed  $S$  rather than  $\text{Tau-b}$  as a measure of alliance portfolio similarity for the risk attitude scores. Signorino and Ritter (1999) provide good reason to believe that  $\text{Tau-b}$  can seriously misrepresent the degree to which two states' alliance portfolios are similar, their measure  $S$  appears to measure better foreign policy similarity.

<sup>16</sup> This "weak-link assumption" is common in quantitative research on the causes of international conflict. See, for example, Dixon 1994 and Oneal and Russett 1997.



and can therefore reflect risk acceptance.<sup>17</sup> The measure could be improved if data on national issue positions and the status quo on those positions were accessible (Morrow 1987, 436). Unfortunately, because these data are not available, we cannot tell whether the failure to distinguish between the security benefits and the autonomy benefits of alliances introduces error into the measure. Despite this shortcoming, Bueno de Mesquita's risk attitude is far from subjective and appears to be a fairly good measure of the concept I wish to operationalize—that states would be relatively more risk-averse in international cooperation with those states they fear the most from a security standpoint.

To create a second risk attitude variable, I use Gartzke and Jo's "Affinity of Nations Index" (2002), which measures preference similarity among states. Preference similarity does not measure directly the level of risk aversion. Still, it is reasonable to assume that states would be relatively more risk-averse with respect to unanticipated changes in the distribution of gains when dealing with partners with very different preferences; in contrast, if a partner with similar preferences gained more, it can be more safely assumed that the gains will be used for shared priorities. That is, it is likely that state A will be increasingly concerned about the distribution of gains with state B, the more divergent their preferences are.

The Affinity indicator reflects the similarity of state preferences based on their voting positions in the United Nations General Assembly (Gartzke and Jo 2002) and is calculated using Signorino and Ritter's (1999) "S" procedure. The possible values range from -1 (least similar interests) to 1 (most similar interests). For ease of interpretation, I invert this scale so that -1 indicates the least risk-averse state and +1 indicates the most risk-averse state. I use Gartzke and Jo's interpolated SUN2CATI variable, which indicates whether a state voted "yes" or "no." Because the Affinity data are dyadic, I simply take the Affinity value for each bilateral agreement. For the multilateral agreements, I first create a dyad for each pair of signatories. Hence, if there are three signatories, there are three dyads; if there are four signatories, there are six dyads, and so on. I then use the Affinity value of the dyad with the least similar interests, utilizing the "weakest link assumption."

The Affinity data's goal of measuring the similarity of state preferences is far from unproblematic, as Gartzke and Jo (2002, 1-2) acknowledge. The key difficulty is that preferences are not directly observable; hence, we must assume that a state's behavior reveals its preferences. The Affinity measure does have two considerable advantages over Bueno de Mesquita's (1985) risk attitude measure. First, it is based on an information source that is less distorted than alliance portfolios. Al-

liances are costly; hence, states that do not have strong motives or threats may not pursue one, even though their preference may be to have an alliance (Gartzke and Jo, 2). Second, there is much more variation in UN voting behavior than in alliance formation, particularly during the Cold War.

The third proxy for risk-aversion is a signatory's GDP growth. I argue that states are least risk-averse when they have either very low or very high growth levels and most risk-averse when growth levels lie in a middle range. The relationship between GDP growth and risk aversion is therefore assumed to follow an inverted-U shape. The motivation for the U-shaped function stems from the literature on electoral incentives and risk attitude and, in particular from those who argue that diversionary foreign policy will occur in times of trouble (Downs and Rocke 1995; Levy 1988; Smith 1996, 1998). The argument is that leaders with domestic problems who anticipate being removed often undertake adventurous foreign policies that they would not have attempted otherwise. It is also widely argued that economic variables, like growth rates, are good predictors of election results. Thus, leaders facing either very low or very high growth rates are likely to be relatively more risk-acceptant (they are either gambling for resurrection or extremely secure) than those with middle levels; the latter do not wish to rock the boat.<sup>18</sup>

I used the Penn World Table and calculated each signatory's change in real gross domestic income (adjusting for terms of trade changes) from the year prior to signing. I again used the weakest link assumption, taking the value of the signatory with the lowest growth rate. Next, I separated the growth rate variable into three equal categories, based on the variable's distribution. States with growth levels from the 0 to 33rd percentile of the distribution were considered low-growth states, states with growth levels from the 33rd to the 66th percentile of the distribution were considered mid-growth states, and finally states with growth levels from the 67th to the 100th percentile of the distribution were considered high-growth states. I then created a dummy variable to designate risk-averse states. This variable equals one for mid-growth states and zero otherwise.

Although my three risk indicators differ in numerous ways, an iterated principal component factor analysis reveals moderately strong factor loadings, ranging from about .27 to .44.<sup>19</sup> These factor loadings indicate that

<sup>17</sup> As Morrow (1987, 436) points out, according to Bueno de Mesquita's calculations, Hitler's Germany appears slightly risk-averse. The risk attitude "assigns to German security the autonomy benefits that Germany derived from its alliances with other revisionist powers."

<sup>18</sup> It is important to note that two thirds of the agreements in the sample are characterized by at least one party that has election data available for the years surrounding the signature date and that missing data does not imply the party was not having elections (e.g., Canada has no data for 1946). As a further check on the appropriateness of this measure for the majority of the agreement observations, I examined how many agreements are characterized by at least one democratic signatory (weakest link assumption). Depending on the cutoff used from the Polity data (5, 6, or 7), at least 142 out of 149 agreements have one democratic signatory. This in itself is an interesting observation worthy of future research. What kinds of states do and do not register their agreements with the United Nations?

<sup>19</sup> The Bueno de Mesquita (1985) and Gartzke and Jo (2002) measures are correlated at .121, the Bueno de Mesquita and growth

**TABLE 1. Intended Duration of Agreements**

Issue Area	Finite or Indefinite? ( <i>p</i> -Value from Test of Independence: 0.003)	
	Percent Finite	Percent Indefinite
Economics	79.7	21.3
Environment	60.0	40.0
Human Rights	44.0	56.0
Security	48.0	52.0
All Agreements	66.4	33.6
Mean Duration of Finite Agreements ( <i>p</i> -Value from Test of Equality of Means: 0.107)		
Issue Area	Mean Duration (Standard Deviation of Duration)	
Economics	12.1	(11.6)
Environment	6.6	(3.5)
Human Rights	8.4	(3.2)
Security	5.0	(0.0)
All Agreements	10.0	(9.4)

*Source:* Author's calculations using data on international agreements. The total sample contains 97 finite agreements and 49 indefinite agreements. Within the economics issue area, the sample includes 19 financial agreements, 17 investment agreements, and 9 monetary agreements.

the three indicators, although not substitutes, are significantly related to one another.

### Duration Patterns in International Agreements

Perhaps the most striking feature of Table 1 (top panel), which presents the intended duration of my random sample of agreements, is that there is a lot of this kind of flexibility: about two thirds of the agreements have a finite duration. As the bottom panel of Table 1 illustrates, the average duration for a finite agreement is about 10 years.

In the top panel of Table 1, the *p*-value of 0.000 indicates that whether an agreement is finite or not depends strongly on issue area. For example, agreements in the economics issue area are almost twice as likely as human rights agreements to be of finite duration. That this kind of flexibility varies in important ways across the broad issue areas suggests that states design their agreements to match the broad features of different issue areas. But flexibility also varies in important ways within issue areas, which suggests that states also tailor their agreements to meet their individual preferences and to reflect the unique aspects of particular agreement contexts. The bottom panel of Table 1 displays

the mean and standard deviation of the intended durations of the finite duration agreements.<sup>20</sup> Again, we see differences across issue area, although they are not as significant as those in the top panel. Of course, the sample size is also smaller, given that I am conditioning on an agreement being finite.

### Subjecting the Model to Empirical Test

Using the absence or presence of a finite (flexible) agreement as the dependent variable, I conduct three probit analyses. In each specification, I include the variable indicating the number of participants (logged) and the uncertainty measure as well as the dummies designating the human rights, economics, and environmental issue areas, as discussed earlier. I test three separate models—one for each measure of risk aversion. Table 2 displays the results of these analyses, and Table 3 shows predicted probabilities.<sup>21</sup>

The statistical analyses provide strong support for the comparative statics. There is robust evidence that as renegotiation costs increase, states become less likely to choose finite agreements (CS-1).<sup>22</sup> Indeed, in all three models, an increase in the number of participants (logged) significantly decreases the probability of a finite agreement ( $p < .001$ ).<sup>23</sup> All else equal, agreements

<sup>20</sup> I exclude those finite duration agreements whose duration is contingent on something outside of the agreement ( $N=36$ ). For example, the durations of some agreements are contingent on domestic laws or other international agreements.

<sup>21</sup> Scrutiny of the agreements in my sample suggests that certain states that conclude a large number of agreements, such as the United States, may follow a template when concluding similar agreements with different states. Thus, it could be argued that I should not be treating my agreements as entirely independent observations. To deal with this potential problem, I ran all three models using robust standard errors. The results do not differ significantly from those displayed in Table 2.

<sup>22</sup> I also operationalize the renegotiation cost variable in an additional way. Some agreements contain language requiring official ratification by the governments of the participating states, others do not. If we were to assume ratification adds to negotiation and renegotiation costs, agreements requiring ratification are more likely to be of indefinite duration and less likely to take the form of a series of renegotiated finite duration agreements (CS-1). I find that not having the ratification requirement significantly increases the probability of a finite agreement. Although this test lends support to my hypothesis, the measure may be problematic. If decisions about which kinds of agreements do or do not need ratification are dictated by customary international law, we can treat these requirements as exogenous to any given agreement. But if states make ratification decisions at the same time they choose flexibility provisions, the test described here suffers endogeneity problems. A further complication arises from the fact that the specific requirements of ratification differ both across states and between the international law point of view and the domestic one. Still, the measure does capture some aspect of renegotiation costs with positive results.

<sup>23</sup> It could perhaps be argued that, as the number of signatories increase, renegotiation becomes less important because the probability of enforcement decreases. On the contrary, according to rational design principles (see Koremenos, Lipson, and Snidal 2001: 789–90), states in these circumstances are more likely to delegate enforcement powers to some centralized authority. This conjecture receives some weak empirical support in Koremenos 2005, who finds that, as the number increases, states are more likely to delegate dispute resolution to third parties ( $p = 0.15$  with an  $n$  of 83).

measures, at  $-.072$ ; and the Gartzke and Jo and growth measures, at  $-.144$ .

**TABLE 2. Results of Probit Analyses of the Presence of a Finite Agreement**

Independent Variable	Model 1 Coefficient (Standard Errors)	Model 2 Coefficient (Standard Errors)	Model 3 Coefficient (Standard Errors)
Number of Participants (Logged)	-.469 (.146)***	-.658 (.201)***	-.323 (.142)**
Uncertainty	1.268 (.279)***	1.223 (.294)***	1.266 (.281)***
Risk Aversion (Bueno de Mesquita 1985)	.648 (.275)**	—	—
Risk Aversion (Gartzke and Jo 2002)	—	.373 (.321)	—
Risk Aversion (GDP Growth)	—	—	.662 (.388)**
Human Rights Issue Area	.848 (.517)	.996 (.548)*	.638 (.512)
Environmental Issue Area	.975 (.450)**	.949 (.457)**	.813 (.436)*
Economics Issue Area	1.110 (.381)***	.865 (.378)**	.831 (.367)**
Constant	-.698 (.421)*	-.116 (.513)	-.664 (.422)
$\chi^2$ , Wald Test of Joint Significance of Number of Participants, Uncertainty, Risk aversion	12.57***	6.67***	9.08***
Percentage Correctly Predicted	.752	.748	.746
Number of Observations	145	135	138

Source: Author's calculations using data on international agreements. \* $p < .10$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ . The Bueno de Mesquita and Gartzke and Jo risk-aversion variables are continuous, and range from -1 (least risk-averse) to +1 (most risk-averse). The GDP growth risk variable is dichotomous and equals 1 for states that have midlevels of growth (and are therefore believed to be most risk-averse). The cutoff point for the calculation of percentage correctly predicted is .664, the sample mean of the dependent variable.

**TABLE 3. Predicted Probability of the Presence of a Finite Agreement Based on Changes in Independent Variables of Interest**

	Predicted Probability at Minimum Value of Independent Variable	Predicted Probability at Mean Value of Independent Variable	Predicted Probability at Maximum Value of Independent Variable
Number of Participants (Model 1)	.800	.706	.169
Uncertainty (Model 1)	.389	—	.832
Risk Aversion (Bueno de Mesquita 1985; Model 1)	.391	.703	.831
Risk Aversion (Gartzke and Jo 2002; Model 2)	.582	.687	.823
Risk Aversion (GDP Growth; Model 3)	.654	—	.844

Note: Predicted probabilities are based on results displayed in Table 2 and were calculated using *Clarify*. All other independent variables were held constant at their mean values.

with only two participants are between 47.5% and 73.7% more likely to be finite (depending on the operationalization of risk aversion employed) than are agreements with the highest number of participants in the sample (119).<sup>24</sup> The results also provide support for the argument that uncertainty increases states' propensity to choose a finite agreement (CS-3). In all three model specifications, the uncertainty effect is in the expected direction and highly statistically significant ( $p < .001$ ). All else equal, agreements with high levels of uncertainty are about 44.0% more likely to be finite than are agreements with low levels of uncertainty.

All three operationalizations provide evidence that risk aversion increases a state's propensity to sign a finite agreement, yielding support for CS-5. The Bueno de Mesquita (1985) risk variable has a statistically significant effect at conventional levels ( $p < .05$ ). According to this model, the most risk-averse participants are 44.0% more likely to create finite agreements than are the least risk-averse participants, all else equal. The

growth risk variable also has a statistically significant effect at  $p < .10$ , indicating that risk-averse participants are 19.0% more likely to create finite agreements than their less risk-averse counterparts, all else equal. The sign of the Affinity risk variable is consistent with the theory, and its magnitude suggests important effects. However, perhaps due to the slightly smaller sample size, this operationalization of risk does not attain statistical significance ( $p = 0.245$ ).<sup>25</sup> While none of the risk measures is ideal, each captures some aspect of what the concept means at the level of states operating in anarchy. Moreover, two of the measures are relational, that is, a state's level of risk aversion depends on its particular partner; the other (growth) is an individual attribute of the state that does not vary with partnership. Whereas international relations scholars are fond

<sup>24</sup> I used *Clarify* (Tomz, Wittenberg, and King 2001) to generate all predicted probabilities, varying the independent variable of interest and holding all other independent variables at their mean values.

<sup>25</sup> An argument can also be made that states become less risk-averse as their citizens become richer. This would imply that states with higher GDPs per capita should be more likely to sign an indefinite agreement. In an alternate specification, I used GDP per capita (taken from the Penn Table) as a proxy for risk-aversion. The results provided some evidence that states with higher GDPs per capita are more likely to sign indefinite agreements, although this result was not significant at standard levels ( $p = .317$ ).

of relational measures, the other measure I have created is closer to the economist's conception.

As a final test of the appropriateness of the three variables, I perform three Wald tests of the null hypothesis that their *joint* effect equals zero.<sup>26</sup> The Wald tests yield strong support for my arguments: the joint effect of renegotiation costs, uncertainty, and risk is significantly different from zero at  $p < .005$  when the Bueno de Mesquita (1985) or growth risk measures are used, and at  $p < .01$  when the Affinity risk measure is employed.

Hence, the renegotiation costs, uncertainty, and risk aversion variables are performing as expected. With regard to the other variables in the analyses, the following findings are of note. In all three models, agreements in both the economics and environmental issue areas are significantly more likely to be of finite duration than are security agreements ( $p$  is between .004 and .062, depending on the variable and the model specification), but they do not differ significantly from those in the human rights area. Human rights agreements are more likely to be finite than security agreements, but the differences are not uniformly statistically significant at conventional levels across the three model specifications.

In an alternative specification, I added four controls to the models: a superpower variable, equal to 1 if the United States or the Soviet Union is a signatory and 0 otherwise; a democracy variable, equal to the mean Polity value of all signatories; a variable indicating whether the agreement contained an intergovernmental provision as a flexibility device; and the signatories' mean GDP per capita. The superpower, intergovernmental provision, and GDP per capita variables consistently bear no relationship to an agreement's intended duration. The democracy variable has a positive and significant effect ( $p < .10$ ) in Model 3, providing some evidence that democracies tend to choose indefinite agreements.<sup>27</sup> In all specifications, the original results remain unchanged by the inclusion of these new variables. A Wald test of the joint significance of the independent variables of central interest—renegotiation costs, uncertainty, and risk aversion—remains significant ( $p < .01$  in two models;  $p < .10$  in one model), whereas a Wald test of the joint significance of the added variables never approaches significance.

My model also predicts that if states choose to conclude a series of renegotiated agreements, increases

in renegotiation costs will lead them to choose agreements with longer intended durations (CS-2); conversely, increases in uncertainty and increases in relative risk aversion will lead states to choose agreements with a shorter duration (CS-4 and CS-6, respectively). To test these predictions, simply performing a regression analysis in which the dependent variable is the intended duration is problematic because the intended duration of some agreements is indefinite. If I limit the sample to those agreements with a finite intended duration, I am selecting on the dependent variable and quite possibly biasing the results. A more appropriate way of testing is to set the intended duration of all indefinite treaties at some value greater than the longest finite treaty. I therefore set the intended duration of all indefinite treaties at four values: 42, 70, 100, and 200 years. Because the data are right-censored, a tobit regression model is used.

I conduct separate tobit analyses using the three risk-aversion proxies. The results provide strong evidence for CS-2. At all four values used, the intended duration of finite agreements increases as renegotiation costs increase ( $p$  is between .001 and .083, depending on the specification). The results also provide strong evidence in support of CS-4. At all four values used, as uncertainty increases, the intended duration of finite agreements decreases notably ( $p < .0001$ ). Finally, at all four values used, the analyses provide some support for CS-6, suggesting that as risk aversion increases, the agreement's intended duration decreases. When the Bueno de Mesquita (1985) variable is used, the result is significant at between .037 and .077. When the Affinity measure is used, the result is significant at between .095 and .283. The coefficient on the growth risk variable is in the expected direction, but falls short of standard levels of statistical significance ( $p$  is between .460 and .806). Table 4 displays the results of the tobit analyses in which indefinite agreements are set at 100. Consistent with the previous analysis, agreements in the security issue area are significantly more likely to be longer than agreements in the other issue areas, suggesting that the security issue area is correlated with some variable(s) not included in my model that also affect agreement length.

Wald tests on Models 4 through 6 provide considerable evidence that renegotiation costs, uncertainty, and risk *jointly* have a significant impact on the intended duration of agreements. The three variables' combined effect is statistically different from zero at  $p < .005$  with all three operationalizations of risk aversion. These results add further and significant support for CS-2, CS-4, and CS-6.

## OTHER FORMS OF FLEXIBILITY

I have shown the choice of duration is affected by uncertainty, renegotiation costs, and relative risk aversion, but is there anything special about duration? We know that other forms of flexibility exist, and there has even been theoretical work on escape clauses.

<sup>26</sup> This is a test of the joint null that all three variables have zero coefficients, rather than a test of each of them one by one. If two (or more) of a model's theoretical variables are highly correlated, they might not show up statistically significant individually due to multicollinearity, but they would show up as statistically significant jointly. I use a two-tailed test, which is conservative given that my alternative hypotheses are in fact one-sided.

<sup>27</sup> Of course, many would argue that Polity is not a good measure of democracy. Critics smartly point out that not only is there no consensus about what should be included (e.g., there is very little overlap among Polity, Freedom House, and Dahl's dimensions), but also there is no reason to assume that changes in one dimension (chief executive constraints) are equivalent to changes in another (competitiveness of political participation). See Vreeland 2003 and those cited therein for a hard-hitting critique.

**TABLE 4. Results of TOBIT Analyses of the Intended Duration of Finite Agreements**

Independent Variable	Model 4 Coefficient (Standard Errors)	Model 5 Coefficient (Standard Errors)	Model 6 Coefficient (Standard Errors)
Number of Participants (logged)	17.218 (6.852)**	32.549 (9.921)***	12.761 (7.255)*
Uncertainty	-67.585 (15.063)***	-68.525 (16.195)***	-72.893 (16.409)***
Risk Aversion (Bueno de Mesquita 1985)	-26.880 (12.814)**	—	—
Risk Aversion (Gartzke and Jo 2002)	—	-25.884 (15.582)*	—
Risk Aversion (GDP growth)	—	—	-12.720 (17.151)
Human Rights Issue Area	-81.032 (29.897)***	-94.319 (31.796)***	-79.713 (31.680)**
Environmental Issue Area	-102.753 (28.014)***	-109.587 (29.350)***	-103.949 (29.546)***
Economics Issue Area	-82.442 (25.875)***	-75.365 (26.728)***	-75.300 (27.498)***
Constant	-175.172 (29.425)***	144.685 (32.464)***	178.186 (31.178)***
$\chi^2$ , Wald Test of Joint Significance of Number of Participants, Uncertainty, Risk aversion	14.61***	10.29***	8.29***
Number of Observations	110 (61 Uncensored)	102 (57 Uncensored)	104 (50 Censored)

Source: Author's calculations using data on international agreements. \* $p < .10$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ . The Bueno de Mesquita and Gartzke and Jo risk-aversion variables are continuous, and range from -1 (least risk-averse) to +1 (most risk-averse). The GDP growth risk variable is dichotomous and equals 1 for states that have midlevels of growth (and are therefore believed to be most risk-averse). Indefinite agreements are set at an intended duration of 100 years.

**TABLE 5. Results of Probit Analyses of the Presence of an Escape Clause**

Independent Variable	Model 1 Coefficient (Standard Errors)	Model 2 Coefficient (Standard Errors)	Model 3 Coefficient (Standard Errors)
Number of Participants (Logged)	.342 (.160)**	.470 (.238)*	.353 (.168)**
Uncertainty	.264 (.425)	.261 (.451)	.182 (.442)
Risk Aversion (Bueno de Mesquita 1985 measure)	-.006 (.450)	—	—
Risk Aversion (Gartzke and Jo 2002 measure)	—	-.197 (.452)	—
Risk Aversion (GDP growth measure)	—	—	-.070 (.592)
Security Issue Area	(Omitted Category)	(Omitted Category)	(Omitted Category)
Human Rights Issue Area	.874 (.563)	1.048 (.597)*	.917 (.578)
Environmental Issue Area	.046 (.709)	.157 (.753)	.074 (.723)
Economics Issue Area	.266 (.599)	.471 (.666)	.095 (.634)
Constant	-2.578 (.680)***	-2.923 (.903)***	-2.542 (.710)***
$\chi^2$ , Wald Test of Joint Significance of Number of Participants, Uncertainty, Risk aversion	.90	.94	.34
Number of Observations	145	135	138

Source: Author's calculations using data on international agreements. \* $p < .10$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ . The Bueno de Mesquita and Gartzke and Jo risk-aversion variables are continuous, and range from -1 (least risk-averse) to +1 (most risk-averse). The GDP growth risk variable is dichotomous, and equals 1 for states that have midlevels of growth (and are therefore believed to be most risk-averse).

To test whether other forms of flexibility can substitute for duration, I perform three probit analyses using the presence of an escape clause as the dependent variable and three using the presence of a withdrawal clause as the dependent variable. The independent variables are identical to those displayed in Table 2. As Tables 5 and 6 demonstrate, most of the independent variables do a very poor job of predicting the outcomes of interest. Uncertainty and risk aversion bear no significant relationship to the probability that an agreement will contain an escape or a withdrawal clause. The effect of the number of participants in each model is not in the expected direction, suggesting that agreements with more participants are more, not less, likely to include escape and withdrawal clauses. Finally, Wald tests that the joint effect of the key variables—number

of participants, uncertainty, and risk aversion—equals zero yield insignificant results across every model.<sup>28</sup> I also analyze a dependent variable coded 1 if an agreement possesses at least one of the aforementioned forms of flexibility (finite duration, escape clause, or withdrawal provision) and 0 otherwise. Not one of the independent variables has a statistically significant impact on flexibility in any of the probits. Together, these results suggest that the determinants of escape and withdrawal provisions are different than those of duration; hence, these other forms of flexibility are solving different problems.

<sup>28</sup> Interestingly, compared to security agreements, human rights agreements are significantly more likely to contain withdrawal clauses across all three models and escape clauses across one model.

**TABLE 6. Results of Probit Analyses of the Presence of a Withdrawal Provision**

Independent Variable	Model 1 Coefficient (Standard Errors)	Model 2 Coefficient (Standard Errors)	Model 3 Coefficient (Standard Errors)
Number of Participants (logged)	.286 (.129)**	.195 (.163)	.253 (.132)*
Uncertainty	.347 (.258)	.271 (.261)	.274 (.260)
Risk Aversion (Bueno de Mesquita 1985 measure)	-.317 (.227)	—	—
Risk Aversion (Gartzke and Jo 2002 measure)	—	.091 (.281)	—
Risk Aversion (GDP growth measure)	—	—	.018 (.286)
Security Issue Area	(Omitted Category)	(Omitted Category)	(Omitted Category)
Human Rights Issue Area	1.090 (.439)**	.993 (.445)**	1.066 (.453)**
Environmental Issue Area	.617 (.377)	.545 (.379)	.653 (.382)*
Economics Issue Area	.466 (.314)	.418 (.323)	.536 (.318)*
Constant	-.706 (.398)*	-.559 (.472)	-.781 (.404)*
$\chi^2$ , Wald Test of Joint Significance of Number of Participants, Uncertainty, Risk Aversion	.74	2.42	1.54
Number of Observations	145	135	138

Source. Author's calculations using data on international agreements. \* $p < .10$ , \*\* $p < .05$ , \*\*\* $p < .01$ . The Bueno de Mesquita and Gartzke and Jo risk aversion variables are continuous, and range from -1 (least risk-averse) to +1 (most risk-averse). The GDP growth risk variable is dichotomous and equals 1 for states that have midlevels of growth (and are therefore believed to be most risk-averse).

What is it that makes a finite duration with the possibility of renegotiation uniquely important in this context? This particular institutional design allows *adjustment* in the face of international uncertainty *without dismantling cooperation*. Escape clauses do not allow adjustment; rather, they allow states to temporarily escape cooperation and return to an unadjusted agreement. Escape clauses are, however, appropriate responses to *domestic* uncertainty. States may agree to particular terms of cooperation but then suffer domestic shocks that make these terms politically difficult. What they require is a temporary relief from their obligations. Even the typical wording of escape clauses suggests this purpose: "extraordinary circumstances that jeopardize extreme national interests." Human rights agreements contain significantly more escape clauses than agreements in other issue areas, with the domestic shock usually being civil war.<sup>29</sup> Article 4 of the "International Covenant on Civil and Political Rights" states that "in time of public emergency which threatens the life of the nation... [states] may take measures derogating from their obligations under the [agreement] to the extent strictly required by the exigencies of the situation." If a state takes such measures, it must inform other state parties through the Secretary-General of the UN regarding "the provisions from which it has derogated and of the reasons by which it was actuated."

Withdrawal clauses are also very different from duration provisions because cooperative institutions cease to exist in the bilateral cases (by far, the majority) or the membership changes in a multilateral setting. The latter can be consequential: when North Korea withdrew from the Nuclear Nonproliferation Treaty, the agreement remained intact, yet the implications of the membership change were serious. Withdrawal

clauses are responses to shocks that alter a state's basic interest in cooperation. Although such shocks rarely occur, the risk they impose is great. Thus withdrawal clauses are pervasive, but their use is infrequent and therefore dramatic.<sup>30</sup>

It is also illuminating that 62% of the agreements have withdrawal provisions and about 8% have escape clauses, but the correlation between these variables and duration never exceeds .18. Flexibility provisions are not simply chosen as a set; nor do particular pairs go together. The problems these provisions uniquely solve occur in different combinations depending on the cooperative endeavor. The conclusion to be drawn is that the landscape of international agreements is far from crude.

Although escape and withdrawal clauses seem to solve different problems than finite duration, another design tool allows adjustment in the face of shocks: an agreement embodying a quasi-legislative institution with the power to modify the distribution of gains. I have argued that such a design may be optimal when uncertainty regarding future gains is pervasive, but renegotiation costs are high because of the number of parties involved. In this context, an institution with an amendment provision characterized by majority rule cuts down on adjustment costs relative to a full renegotiation (Koremenos 2000).<sup>31</sup> The International Monetary Fund provides a good example of an indefinite

<sup>29</sup> This fact poses a challenge to those who believe rationalist logic is inappropriate in this issue area.

<sup>30</sup> The relationship between "bedrock" preferences, which are fundamentally stable, and constraints, which arise from the fact that the state is a composite actor, also provides insight (Koremenos, Lipson, and Snidal 2001, 1073). I would argue that withdrawal clauses are used in the event of "bedrock" changes, whereas escape clauses are used in the event of unchanged bedrock preferences but different domestic constraints.

<sup>31</sup> Not all amendment provisions are substitutes for finite duration. Often amendments are binding only on those accepting them. Environmental agreements have the most amendment provisions at 33%.

duration agreement that establishes an institution that does many things including, importantly, adjusting the distribution of gains. As I mentioned earlier, this option is left out of the full analysis because of current data limitations, but future work will address this question of substitutability.<sup>32</sup>

## OTHER THEORIES

Because this analysis fits squarely within the rationalist institutionalist paradigm in international relations, I now consider what the two competing theoretical approaches, constructivism and realism, would predict, although neither has generated testable implications regarding international institutional design. With respect to constructivists, Finnemore's (1996) work draws on sociological institutionalism, which documents isomorphic outcomes across a variety of substantive areas in different parts of the world. Similar outcomes are driven by a common global culture. If we were to take this argument to the level of agreement design, we would expect emulation regarding agreement form. Additionally, Finnemore and Barnett (1999) argue that agency culture drives institutional form, so the World Bank looks like it does because economists run it. Of course, most agreements are bilateral and establish no agency. Still, it could be argued that international lawyers have a culture; because they write the agreements, these lawyers are likely to follow a template. The design implication of using templates is observationally equivalent to the one predicted by isomorphism: emulation.

To capture the idea of emulation manifested through either more or less frequent use of finite duration provisions over time, I ran a test with time as the only independent variable. The test yielded no significant results. Additionally, the descriptive statistics presented previously also suggest that world culture does not produce isomorphism with respect to flexibility provisions; rather, what emerges is striking variation.

Another line of constructivist scholars emphasizes the role of norms and activist networks in shaping international outcomes. However, the primary outcomes being explained are whether states form cooperative institutions or not and, if they do, whose values are reflected in those institutions. These scholars are not explaining agreement design. In fact, Keck and Sikkink (1998) argue that even normative actors need to be strategic and therefore should pay attention to the de-

tails of institutional design. Hence, it could be argued that my analysis complements their analysis because we are predicting outcomes at very different levels.

Realists, who focus on power considerations, do not have a theory of institutions; they have not taken them seriously. (Glaser [1994] is one of a few realists writing about cooperation, but he does not believe that institutions, let alone design, matter.) Hence, variation in flexibility provisions would not surprise them. Importantly, however, they would expect no systematic variation given their argument that international agreements and institutions are epiphenomenal; rather, any variation would be noise. The data clearly refute this prediction. (And, as discussed previously, the superpower variable is insignificant.)<sup>33</sup>

## CONCLUSIONS

The effects of uncertainty are dramatic in international relations given the fundamental insecurity posed by anarchy. How do states transcend this uncertainty and cooperate? The answer is straightforward: they negotiate agreements that include the proper amount of flexibility and thereby create for themselves a kind of international insurance.

Challenging a long and widely held view that anarchy makes international relations qualitatively distinct from other fields, I develop a model of duration provisions based on economic contracting theory that highlights renegotiation costs, uncertainty regarding the distribution of gains, and relative risk aversion. I then subject this model to a test featuring a random sample of international agreements.<sup>34</sup>

The descriptive statistics alone force us to reconsider one of the conventional wisdoms concerning international cooperation: states tie their hands in order to make their commitments credible. In fact, two thirds of the agreements in this sample are designed *not* to last forever. Does this statistic imply that states are not making many credible commitments? On the contrary, my model suggests that flexibility may *enhance* credibility in environments subject to shocks; reneging, which can be quite damaging, can often be avoided if renegotiation is an option instead.

It could be argued that because some of the agreements in the random sample are minor, the conclusions I draw may be meaningless. Yet it is doubtful these agreements are minor to the governments that signed them, and these sorts of agreements govern most of

<sup>32</sup> The degree of precision/ambiguity in an agreement's obligations is another form of flexibility that might be a substitute for duration, although it would not be a good response to the persistent uncertainty modeled in this paper. Rather, imprecision could be a response to a one-time initial uncertainty, like that modeled in Koremenos 2001. The current broader recoding of each of the agreements in the sample tries to capture such a variable with four categories of precision/ambiguity. In future work, the kinds of analyses performed with escape clauses and withdrawal clauses will be possible with the precision variable. See Abbott and Snidal 2000, for a discussion of precision and of soft law more generally.

<sup>33</sup> This discussion reflects a first (yet serious) attempt at teasing out some testable implications from these other paradigms. I welcome others with more expertise to elaborate alternative testable hypotheses.

<sup>34</sup> Duffield (2003, 426) criticizes certain rationalist approaches to international cooperation for their reliance on case studies. Quite rightly he argues that nothing can be "proved" with a case study. This empirical analysis therefore meets the Duffield challenge. Nonetheless, case studies are also essential. They reveal the historical importance of our theoretical propositions, they help us understand the mechanisms by which our causal variables operate, and they provide us with much needed guidance regarding how to operationalize our variables.



the day-to-day cooperation between states. So if cooperation is to be explained, such agreements deserve attention as a category.<sup>35</sup> Certainly, future work must address what happens once the agreements are designed. Although a number of case studies illustrate that duration and renegotiation provisions are meaningful with respect to how cooperation evolves, an important next step would be to randomly select 15 to 20 of the agreements used in this analysis and research and evaluate their effectiveness.

The sharpest potential critique, however, is that it makes no sense to focus analysis on agreement design when the value of international agreements is itself contested. Still, one way of shedding light on this particular international relations debate is to examine whether the design of agreements is consistent with the presumption that the agreements themselves are important. In fact, instead of random variations among agreements or automatic replication of the same agreement provisions over and over, the detailed provisions of international agreements are chosen in ways to increase the incidence and robustness of cooperation. States would not spend so much effort getting institutions "right" if the institutions didn't matter.

International agreements are nuanced and sophisticated—just like the domestic agreements and institutions studied in American and comparative politics and in law and economics.<sup>36</sup> Although anarchy makes uncertainty dramatic, states contract around this uncertainty with various forms and combinations of flexibility, like duration, escape, and withdrawal provisions. Each is a unique response to a distinct kind of uncertainty. Nor is flexibility the only agreement feature that proves systematic. Additional analyses (e.g., Koremenos N.d.) suggest that provisions regarding monitoring and dispute resolution also are chosen according to principles of rational design.

International agreements are important because they regulate cooperation, and the fact that they obey lawlike regularities indicates that serious efforts are made for them to be able to regulate interactions in lasting and successful ways. So, because agreements matter, they are designed in rational ways, and the fact that people make efforts to design them in such ways corroborates their significance.

<sup>35</sup> Of course, our intuitions about the forms of international cooperation are often shaped by a few high-profile agreements. High-profile agreements get that way because they have big effects; thus, it is important to be able to make statements about their characteristics. Future work will build a sample of high-profile agreements, and we will be able to make a comparison between the two samples. Only then can we say with any confidence whether past reliance on case studies was somewhat justified.

<sup>36</sup> International relations is not the only field implicated in this analysis. A critical debate in law and economics centers on the importance of the "shadow of the law." Given that the design of international cooperation obeys lawlike regularities without any compelling supranational authority, perhaps the "law" is not quite as important as many thought *within* the state. Not only can international relations scholars learn from law scholars, but also we have important results to share with them.

## APPENDIX A

Assuming that State A and B are symmetric, compare State A's expected utility with one two-period (inflexible) agreement, given by  $EU_2 = E(u(b_1 + m_0 + e_1 - k_r) + \delta u(b_1 + m_0 + e_1 + e_2))$ , to its expected utility from two 1-period (flexible) agreements, given by  $EU_1 = E(u(b_1 + m_0 + e_1 - k_r) + \delta u(b_1 + m_0 + e_2 - k_r))$ . The only differences between  $EU_1$  and  $EU_2$  occur in the second period. In that period, with  $EU_2$  the states again live with the first period shock  $e_1$  but do not pay renegotiation cost  $k_r$ . The reverse is true for  $EU_1$ , where the states avoid having to experience  $e_1$  by paying the renegotiation cost  $k_r$ .

With respect to CS1, the key point is that  $EU_1$  is decreasing in the renegotiation cost  $k_r$  whereas  $EU_2$  is unaffected by  $k_r$ . Thus, suppose that initially  $EU_2 > EU_1$ ; that is, an inflexible agreement dominates. Holding the distribution of  $e_1$  fixed, we can decrease  $k_r$  until at some point a threshold is reached and  $EU_1 > EU_2$ . States will then change from an inflexible two-period agreement to a flexible series of one-period agreements. A similar story holds in the other direction, where initially two 1-period agreements are preferred. Increasing  $k_r$  will decrease  $EU_1$  until at some point the two expected utilities are equal, and after which point the states jointly prefer an inflexible two-period agreement because the increased cost of renegotiation makes the one-period agreements no longer optimal.

With respect to CS3, holding the form of the utility function constant and assuming that it is strictly increasing and concave, a mean-preserving increase in the variance of the shocks to the distribution of gains (i.e., an increase in uncertainty),  $e$ , reduces both  $EU_1$  and  $EU_2$ , because  $e_1$  and  $e_2$  enter both and because concavity of the utility function implies risk aversion (i.e., states prefer a lower variance to the shocks at a given mean, in this case zero). But because the shocks are independent, in the second period, the combined shock ( $e_1 + e_2$ ) experienced under the two-period agreement has a variance equal to  $2 \text{ var}(e)$ , whereas the second period shock in the case of two 1-period agreements, just  $e_2$ , has only  $\text{var}(e)$ . Thus, a given increase in  $\text{var}(e)$ , holding the mean constant, increases the variance of the combined shock more for the two-period agreement, and thereby leads to a larger decrease in  $EU_2$  than in  $EU_1$ . That is, increasing the variance of the shocks decreases  $EU_2$  relative to  $EU_1$ . Thus, if we start from a situation where an inflexible two-period agreement is preferred by the states and then continually increase the  $\text{var}(e)$ , holding constant  $k_r$  and the mean of the shocks, at some point  $EU_2$  will fall sufficiently relative to  $EU_1$  that the states will prefer a flexible regime of two 1-period agreements. At this point, it is worth it to the states to pay  $k_r$  in order to reduce the variance of the second-period outcomes. The same argument holds in the other direction. If initially the states prefer two 1-period agreements, we can decrease  $\text{var}(e)$ , which will now increase  $EU_2$  relative to  $EU_1$ , until  $EU_2 > EU_1$ . In the limiting case, where  $\text{var}(e) = 0$ , states will prefer a two-period agreement as long as  $k_r > 0$ .

A similar argument holds with respect to CS5. Here we hold the variance of the shocks constant while varying the concavity of the utility function. The key point is that the variance of the outcomes is higher under the inflexible two-period agreement, given that states experience the combined shock ( $e_1 + e_2$ ) in the second period. As a result, increasing the concavity of the utility function (i.e., increasing the degree of risk aversion) reduces both expected utilities,  $EU_1$  and  $EU_2$ , but decreases  $EU_1$  relatively more. Suppose initially the two-period agreement is preferred by both states. If we increase the concavity of the utility function, say by increasing the degree of the exponent,  $\alpha$ , when  $u(x) = x^\alpha$  as in

**TABLE B1. Results from Simulation of Two-Period Model**

Utility Function	$\delta$	$var(\varepsilon)$	$k_n$	$k_r$	$c$	$E(U_R)$	$E(U_N)$	Choice
Base Case								
Square Root	0.9	Low	1.0	0.5	1.0	12.4927	12.5130	One 2-Period (Inflexible)
Decrease Renegotiation Costs								
Square Root	0.9	Low	1.0	0.1	1.0	12.5199	12.5185	Two 1-Period (Flexible)
Increase Variance of Shocks (Uncertainty)								
Square Root	0.9	Hlgh	1.0	0.5	1.0	12.4900	12.4557	Two 1-Period (Flexible)
Increase Risk Aversion								
Cubic	0.9	Low	1.0	0.5	1.0	6.6682	6.6658	Two 1-Period (Flexible)

the simulations, or by increasing the risk aversion parameter  $a$  in a constant absolute risk aversion (CARA) utility function where  $u(x) = -\alpha e^{-\alpha x} + \beta$  (Kreps 1990: 86),  $EU_2$  declines relative to  $EU_1$ . At some point,  $EU_1 > EU_2$ , and both states now prefer to enter into two 1-period agreements. As previously, the same arguments work in reverse if we start from a situation where the states choose two 1-period agreements, and then decrease the level of concavity (and therefore risk aversion) in the utility function. Eventually the states will become risk-tolerant enough that they prefer an inflexible two-period agreement to two 1-period agreements, so long as  $k_r > 0$ . In the limiting case of a linear utility function (risk neutrality), the states will always prefer the longer agreement if  $k_r > 0$ .

## APPENDIX B

This appendix describes the construction of the solutions to the discretized version of the two-period model used for the simulations. The value of base outcome,  $a$ , is set to 40.0. This value is completely arbitrary as long as it is positive; its main function in the simulations is to be large enough to keep the outcome from going negative even after a series of bad agreement shocks. The value of each state's share of the initial distribution of gains is set to  $m=4$ , which implies a total gain from the agreement, given equal *ex ante* division due to Nash bargaining, of  $g=2m=8$ . The general class of utility functions considered in the simulation is polynomial. In particular, I consider  $U(X)=X$ ,  $U(X)=X^2$ ,  $U(X)=X^3$ ,  $U(X)=X^4$  or, in words, a linear utility function, a quadratic utility function a cubic utility function, and a quartic utility function. For the values of  $X$  considered here, which are all positive and greater than 1, an increase in the exponent of the utility function represents an increase in the level of risk aversion.

To keep the programming simple, the distribution of shocks is discrete, not continuous. The shock in the first period is denoted  $\varepsilon_1$ , and the shock in the second period is denoted  $\varepsilon_2$ . The potential values of the shocks to the distribution of gains under the agreement are denoted by  $\varepsilon_{1j}$  and  $\varepsilon_{2j}$ , where  $j \in \{1, 2, 3\}$  denotes one of the three possible values of the shocks. In the simulations in Table B1, the values of the shocks are  $-2, 0$  and  $2$  so that  $\varepsilon_{11} = -2$ , and so on. The variance of the distribution of shocks is increased/decreased through a mean-preserving increase/decrease in the spread of the distribution; in this case, there is a symmetric increase in the probability of the two extreme values, with a corresponding reduction in the probability of a zero shock.

I examined values for the costs of negotiation,  $k_n$ , and renegotiation,  $k_r$ , in the range  $[0, 1]$ . I considered values of the discount rate,  $\delta$ , in the set  $\{0.6, 0.7, 0.8, 0.9\}$ , whereas the values of the costs of reneging,  $c$ , ranged over the interval

$[0, 5]$ . For values of  $c$  near the high end of this range (how high depends on the shape of the utility function), reneging does not occur given the values of the shocks utilized here.

For each of the variables corresponding to the comparative static results, I examined essentially the entire range of possible values (or distributions) within this framework; all were (not surprisingly given the proofs in Appendix A), consistent with the model. The particular values presented in Table B1 were chosen to represent the wide range of simulation results actually generated.

The expected utility associated with two 1-period agreements is given by

$$E(U_R) = \sum_{j=1}^3 \Pr(\varepsilon_1 = \varepsilon_{1j}) U(a + m + \varepsilon_{1j}) \\ + \delta \sum_{j=1}^3 \Pr(\varepsilon_2 = \varepsilon_{2j}) U(a + m + \varepsilon_{2j} - k_r),$$

where  $U(\cdot)$  is the utility function.

The equation for a single two-period agreement is more complicated due to the possibility of reneging. Consider the problem from the standpoint of state A. We can define an indicator variable,  $R_A$ , for whether or not State A will renege as follows:

$$R_A(\varepsilon_{1j}) = 1[E(U(a + m + \varepsilon_2 - k_r - c)) \\ > E(U(a + m + \varepsilon_{1j} + \varepsilon_2))],$$

where the expectations are over the possible values of  $\varepsilon_2$ . The expectations are as of the end of the first period, so that  $\varepsilon_1$  is known but  $\varepsilon_2$  is not. In other words, State A compares the expected utility of reneging, given by the first term, with the expected utility of not reneging, given by the second term. The gain from reneging is getting rid of  $\varepsilon_1$ ; the cost is the combined cost of reneging plus renegotiation, given by  $(k_r + c)$ . State B makes a similar calculation; we can define an indicator variable  $R_B$  for its decision as follows:

$$R_B(\varepsilon_{1j}) = 1[E(U(a + m - \varepsilon_2 - k_r - c)) \\ > E(U(a + m - \varepsilon_{1j} - \varepsilon_2))].$$

With these indicator variables in hand, we can now give the formula for State A's expected utility from a two-period agreement. It is

$$E(U_N) = \sum_{j=1}^3 \Pr(\varepsilon_1 = \varepsilon_{1j}) \{U(a + m + \varepsilon_{1j} - k_n) \\ + \delta[R_A(\varepsilon_{1j})T_1 + R_B(\varepsilon_{1j})T_2 + (1 - R_A(\varepsilon_{1j}) \\ - R_B(\varepsilon_{1j}))T_3]\},$$

where

$$T_1 = \sum_{j=1}^3 \Pr(\varepsilon_2 = \varepsilon_{2j}) U(a + m + \varepsilon_{2j} - c - k_r),$$

$$T_2 = \sum_{j=1}^3 \Pr(\varepsilon_2 = \varepsilon_{2j}) U(a + m + \varepsilon_{2j} - k_r),$$

and

$$T_3 = \sum_{j=1}^3 \Pr(\varepsilon_2 = \varepsilon_{2j}) U(a + m + \varepsilon_{1j} + \varepsilon_{2j}).$$

The formulae for State B are symmetric.

## REFERENCES

- Abbott, Kenneth W., and Duncan Snidal. 2000. "Hard and Soft Law in International Governance." *International Organization* 54 (Summer): 421–56.
- Barnett, Michael N., and Martha Finnemore. 1999. "The Politics, Power, and Pathologies of International Organizations." *International Organization* 53 (Autumn): 699–723.
- Bennett, D. Scott, and Allan Stam. 2000a. "EUGene. A Conceptual Manual." *International Interactions* 26, 179–204.
- Bennett, D. Scott, and Allan C. Stam. 2000b. "A Universal Test of an Expected Utility Theory of War." *International Studies Quarterly* 44 (September): 451–80.
- Bordo, Michael, and Finn Kydland. 1995. "The Gold Standard as a Rule: An Essay in Exploration." *Explorations in Economic History* 32 (October): 423–64.
- Brown, Michael E., ed. 2000. *Rational Choice and Security Studies: Stephen Walt and His Critics*. Cambridge: MIT Press.
- Bueno de Mesquita, Bruce. 1985. "The War Trap Revisited: A Revised Expected Utility Model." *American Political Science Review* 79 (March): 156–77.
- Conybeare, John. 1985. "Trade Wars. A Comparative Study of Anglo-Hanse, Franco-Italian, and Hawley-Smoot Conflicts." *World Politics* 38 (October): 147–72.
- Dixon, William J. 1994. "Democracy and the Peaceful Settlement of International Conflict." *The American Political Science Review* 88 (March): 14–32.
- Downs, George, and David Rocke. 1995. *Optimal Imperfection?: Domestic Uncertainty and Institutions in International Relations*. Princeton, NJ: Princeton University Press.
- Duffield, John. 2003. "The Limits of 'Rational Design.'" *International Organization* 57 (Spring): 411–30.
- Dye, Ronald A. 1985. "Optimal Length of Labor Contracts." *International Economic Review* 26 (February): 233–50.
- Finnemore, Martha. 1996. *National Interests in International Society*. Ithaca, NY: Cornell University Press.
- Gartzke, Erik, and Dong-Joon Jo. 2002. "The Affinity of Nations Index, 1946–1996." Version 3.0; <http://www.columbia.edu/~eg589/datasets.html>, November 8, 2002.
- Glaser, Charles. 1994. "Realists as Optimists: Cooperation as Self Help." *International Security* 19 (Winter): 50–90.
- Gray, Jo Anna. 1978. "On Indexation and Contract Length." *Journal of Political Economy* 86 (February): 1–18.
- Guzman, Andrew, and Beth A. Simmons. 2002. "Litigation and Settlement." *Journal of Legal Studies* 31: S205–S235.
- Harris, Milton, and Bengt Holmstrom. 1987. "On the Duration of Agreements." *International Economic Review* 26 (June): 389–406.
- Keck, Margaret E., and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.
- Koremenos, Barbara. 2000. "Bending but Not Breaking: Flexibility in International Financial and Monetary Agreements." Working Paper 1.73. University of California, Berkeley: Center for German and European Studies.
- Koremenos, Barbara. 2001. "Loosening the Ties that Bind: A Learning Model of Agreement Flexibility." *International Organization* 55 (Spring): 289–325.
- Koremenos, Barbara. 2005. "Bringing More 'Precision' to the Three Dimensions of Legalization." Paper presented at the European University Institute, Fiesole, Italy.
- Koremenos, Barbara. N.d. "If Only Half of International Agreements Have Dispute Resolution Provisions, Which Half Needs Explaining?" *Journal of Legal Studies*. Forthcoming.
- Koremenos, Barbara. <http://www.polisci.ucla.edu/faculty/koremenos>
- Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. "Rational Design: Looking Back to Move Forward." *International Organization* 55 (Autumn): 1051–82.
- Kreps, David M. 1990. *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.
- Levy, Jack S. 1988. "When do Deterrent Threats Work?" *British Journal of Political Science* 18 (October): 485–512.
- Martin, Lisa L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.
- Martin, Lisa L. 2000. *Democratic Commitments: Legislatures and International Cooperation*. Princeton, NJ: Princeton University Press.
- Morrow, James D. 1987. "On the Theoretical Basis of a Measure of National Risk Attitudes." *International Studies Quarterly* 31 (December): 423–38.
- Oneal, John R., and Bruce M. Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41 (June): 267–93.
- O'Neill, Barry. 2001. "Risk Aversion in International Relations Theory." *International Studies Quarterly* 45 (December): 617–40.
- Osborne, Martin J., and Ariel Rubinstein. 1990. *Bargaining and Markets*. Boston: Academic Press.
- Rosendorff, Peter, and Helen Milner. 2001. "The Optimal Design of International Trade Institutions: The Sources and Impact of Escape Clauses." In *The Rational Design of International Institutions*, ed. Barbara Koremenos, Charles Lipson, and Duncan Snidal. *International Organization*. Special Issue 55 (Autumn): 829–57.
- Signorino, Curtis, and Jeffrey Ritter. 1999. "Tau-b or Not Tau-b. Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly* 43 (March): 115–44.
- Simmons, Beth A. 2000. "International Law and State Behavior. Compliance with the Public International Law of Money." *American Political Science Review*, 94 (4): 819–35.
- Smith, Alastair. 1996. "Diversionary Foreign Policy in Democratic Systems." *International Studies Quarterly* 40 (March): 133–53.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (September): 623–38.
- Tomz, Michael, Jason Wittenberg, and Gary King. 2001. "Clarify: Software for Interpreting and Presenting Statistical Results." <http://gking.harvard.edu/clarify> (April 23, 2003).
- Vreeland, James Raymond. 2003. "A Continuous Schumpeterian Conception of Democracy." Working Paper, Yale University, New Haven, CT.

# Centripetal Democratic Governance: A Theory and Global Inquiry

JOHN GERRING *Boston University*

STROM C. THACKER *Boston University*

CAROLA MORENO *Boston University*

**W**hy are some democratic governments more successful than others? What impact do various political institutions have on the quality of governance? This paper develops and tests a new theory of democratic governance. This theory, which we label centripetalism, stands in contrast to the dominant paradigm of decentralism. The centripetal theory of governance argues that democratic institutions work best when they are able to reconcile the twin goals of centralized authority and broad inclusion. At the constitutional level, our theory argues that unitary, parliamentary, and list-PR systems (as opposed to decentralized federal, presidential, and nonproportional ones) help promote both authority and inclusion, and therefore better governance outcomes. We test the theory by examining the impact of centripetalism on eight indicators of governance that range across the areas of state capacity, economic policy and performance, and human development. Results are consistent with the theory and robust to a variety of specifications.

**W**hy are some democracies better governed than others? Why are many plagued by corruption and ineptitude, whereas others manage to implement policies effectively and efficiently? Why are some borne down by inefficient markets and low standards of living, whereas others enjoy low transaction costs, high capital investment, and strong economic performance? Why are rates of morbidity, mortality, illiteracy, and other aspects of human deprivation so depressingly high in some democracies, and so impressively low in others? What can account, in short, for the immense variation we observe in the quality of governance across democratic polities in the world today?

In this paper, we focus on the role of democratic political institutions in the achievement of good governance. The survival of democracy is understood as

a peripheral—albeit important—question (Linz 1994; Stepan and Skach 1993). Presumably, the quality of governance influences the propensity of a democracy to survive; however, we do not theorize this question. We understand a country to be democratic when multiparty competition, under reasonably fair conditions, is in place. We are specifically concerned with the role of political institutions in achieving good governance. Other factors—geographic, economic, historical, sociological, or cultural—lie in the background.

Two opposing perspectives on this question have predominated since the advent of representative government in the eighteenth century. We label these primordial theories *centralism* and *decentralism*. The centralist theory, closely associated with the Westminster system and the theory of Responsible Party Government, presumes that good governance flows from institutions that centralize power in a single locus of sovereignty. The decentralist theory, associated with the American polity and with a variety of theoretical frameworks, supposes that good governance arises from the diffusion of power among multiple independent bodies. Simply formulated, the governance debate over the past two centuries has been an argument between Hobbes and Montesquieu.

More recently, the Hobbesian model seems to have lost much of its vigor and appeal. Scholars today rarely appeal to the virtues of Westminster. Accordingly, there are few democratic centralists at the present time, either in the academy or in the world of policymaking and politics. Both the Left and the Right now apparently agree on the virtues of decentralized democratic institutions.

Our intention in this paper is to present a revived, and significantly modified, version of democratic centralism. We argue that democratic institutions work best when they are able to reconcile two goals: centralized authority and broad inclusion. Good governance should arise when political institutions preserve the authority of the sovereign while gathering together and effectively representing whatever ideas, interests, and identities are extant in a society. These

John Gerring is Professor, Department of Political Science, Boston University, 232 Bay State Road, Boston MA 02215 (jgerring@bu.edu).

Strom C. Thacker is Professor, Boston University, Department of International Relations, 152 Bay State Road, Boston, MA 02215 (sthacker@bu.edu).

Carola Moreno is Ph.D. candidate, Department of Economics, 270 Bay State Road, Boston, MA 02215 (cmoreno@bu.edu).

We are grateful for comments and suggestions received from participants in seminars at Brown University, Cornell University, Duke University, Princeton University, and the University of Leuven-la-Neuve (Belgium), where various versions of this paper were presented. For shared data, feedback, and good counsel we would like to acknowledge Victor Aguirregabiria, Pablo Beramendi, Tom Burke, John Carey, Lieven De Winter, Kent Eaton, Neil Englehart, Macartan Humphreys, Mark Jones, Herbert Kitschelt, Atul Kohli, Chappell Lawson, Evan Lieberman, Arend Lijphart, Branko Milanovic, Howard Reiter, Benoit Rihoux, David Samuels, Matthew Shugart, Nicolas van de Walle, David Waldner, Robert Weiner, Christopher Whytock, and Deborah Yashar. Joshua Yesnowitz provided critical research assistance at the final stages. We are also grateful to three anonymous reviewers who provided helpful comments and suggestions. The standard disclaimers apply. In addition, we would like to acknowledge the generous financial support that we have received from the Institute for Advanced Study (Gerring), the Hoover Institution (Thacker), and the Frederick S. Pardee Center for the Study of the Longer-Range Future at Boston University.

twin goals are captured in the concept of *centripetalism*, which we employ as a label for this new theory of governance.<sup>1</sup>

Empirically, we expect the theory of centripetalism to operate at multiple levels—local, regional, national, and international. In this paper, our vision is restricted to the national level and to three constitutional institutions that, we feel, best embody the centripetal ideal: unitarism, parliamentarism, and a closed-list PR electoral system. These are the building blocks of the centripetal polity (when that polity is democratic) and the centerpiece of our empirical investigation. To measure good governance outcomes, we employ a battery of indicators focused on various facets of political, economic, and human development. We regress these indicators against our principal theoretical variable—a composite measure of unitarism, parliamentarism, and list-PR—in a global sample of democratic polities. Such tests provide support for the hypothesis that political institutions fostering centralized authority and broad inclusion lead to better governance.

Although the precise causal mechanisms at work in the relationship between centripetal institutions and good governance are difficult to specify and to measure—and therefore virtually impossible to test—we speculate that centripetal institutions encourage strong political parties, corporatist-style interest representation, collegial decisionmaking, and authoritative public administration. Each of these intermediate factors should foster better governance in democratic polities. We therefore regard each one as an important causal pathway in our macrotheoretical argument.

## DECENTRALISM

The decentralist model of governance that predominates among contemporary scholars and policymakers emerged from a centuries-long struggle for political accountability in the West. This history begins with the classical polities of Greece and Rome and continues through the British, Italian, Swiss, and Dutch polities of the early modern era (Gordon 1999; Vile 1967/1998). Thus, by the time of the American Revolution, the essential features of this model of democratic governance were already in place. Hereafter, the American polity came to be viewed as the paragon of decentralism, and the Federalist Papers as its interpretive catechism.

Among twentieth-century writers, decentralism takes a number of different forms, each with its own terminology, theoretical framework, and policy concerns. This far-ranging camp includes early group theorists (Bentley 1908/1967), British pluralists (Hirst 1989), American pluralists (Dahl 1956; Herring 1940; Truman 1951), Guillermo O'Donnell's (1999) conception of horizontal accountability, Arend Lijphart's (1999) consensus model of governance, and various writers in the public choice tradition, especially as oriented around the intertwined ideas of separate powers, fiscal

federalism, veto points, and insulation (Buchanan and Tullock 1962; Henisz 2000; North and Weingast 1989; Oates 1972; Persson, Roland, and Tabellini 1997; Tiebout 1956). Decentralism is a broad church with many followers.

A central division among decentralists concerns attitudes toward popular rule. The dominant strand, beginning with Blackstone, Montesquieu, and Madison, sees in decentralization a mechanism to prevent direct popular rule, or at least to moderate its effects. A majoritarian system, it is feared, is prey to manipulation by unscrupulous leaders and envious masses bent on the redistribution of wealth (e.g., Riker 1982). An opposing strand, associated with Paine, Rousseau, and others of a radical or populist persuasion, perceives the decentralization of power as a mechanism to bring government closer to the people. Their assumption is that centralized power is generally controlled by leaders whose interests run contrary to the electorate; the only hope for popular control of government is therefore to decentralize the locus of decisionmaking. Thus, the wellsprings of decentralism lie in suspicions of elites and/or of the masses.

Despite their evident differences, all twentieth-century decentralists agree with several core precepts: diffusion of power, broad political participation, and limits on governmental action. Separate powers and federalism are the two key theoretical components; one implies divisions on a horizontal dimension; the other, on a vertical dimension. Institutional fragmentation at both levels is intended to set barriers against the abuse of power by minorities, against the overweening ambitions of individual leaders, against democratic tyrannies instituted by the majority, and against hasty and ill-considered public policies. Decentralist government is limited government. Each independent institution is intended to act as a check against the others, establishing a high level of interbranch accountability. Bad laws have little chance of enactment in a system biased heavily against change, where multiple groups possess an effective veto power over public policy. The existence of multiple veto points forces a consensual style of decisionmaking in which all organized groups are compelled to reach agreement on matters affecting the polity. Limitations on central state authority preserve the strength and autonomy of the market and of civil society, which are viewed as separate and independent spheres (as emphasized by those in the Madisonian camp). Decentralized authority structures may also lead to greater popular control of, and direct participation in, political decisionmaking (as emphasized by the Rousseauian camp). Efficiency is enhanced by political bodies that lie close to the constituents they serve, by a flexible apparatus that adjusts to local and regional differences, and through competition that is set into motion among semiautonomous governmental units.

How do these theoretical desiderata translate into specific political institutions? The principle of separate powers suggests two elective lawmaking authorities as well as a strong and independent judiciary. The principle of federalism presumes a shared sovereignty

<sup>1</sup> The term *centripetalism* has been employed in the context of state-building (Bryce 1905) and party competition (Cox 1990; Sartori 1976), but not in the context of overall governance.

composed of national and subnational units. Both also suggest a bicameral legislature—to further divide power at the apex and to ensure regional representation. In addition, the decentralist model implies a written constitution, perhaps with enumerated individual rights and explicit restrictions on the authority of the central state, and strong local government. Most decentralists embrace the single-member district as a principle of electoral law that maximizes local-level accountability. There is disagreement over whether this should be supplemented by mechanisms to enhance intraparty democracy, for example, open primaries or preferential-vote options. If we take the principle of decentralism literally, we are led toward several additional institutional features: numerous elective offices, frequent elections (short terms), staggered terms of office, nonconcurrent elections, fixed-term elections (no possibility of premature dissolution), term limits, popular referenda, recall elections, decentralized party structures, agencies enjoying a high degree of independence, and small political units (micro- rather than macro states).

Although one might quibble over details, there is consensus on the basic institutional embodiments of a decentralist political order, where power is diffused among multiple independent actors and energy flows in a centrifugal direction toward the peripheries. This is the reigning paradigm of good governance in academic and policymaking circles at the turn of the twenty-first century.

## CENTRIPETALISM

In contrast to the theory of decentralism, we propose that good government results when political energies are focused toward the center. Centripetal, rather than centrifugal, institutions create the conditions for good governance. This idea also has deep historical roots in the Anglo-European tradition. Progenitors include Bodin and Hobbes, who developed the modern concept of sovereignty. In the democratic era, the theory of centripetalism may be understood as a melding of two distinct theories of governance, the Responsible Party Government (RPG) model and the less clearly defined model of governance elaborated by early proportional representation (PR) reformers.

The RPG model, beginning with Walter Bagehot and Woodrow Wilson and extending to later work by E. E. Schattschneider and many others, is a model of democratic centralism (Ranney 1962). This vision of politics also informs work by defenders of the welfare state and of strong government in the contemporary era, who see multiple veto points as the source of special-interest pressures (e.g., Lowi 1969; McConnell 1966). For this diverse group, comprised mostly of social democrats, that system is best which focuses all power on a single locus of sovereignty: the prime minister and his cabinet. Party control of the legislature allows for a temporary dictatorship; mechanisms of electoral accountability ensure that this period of one-party rule will be in the public interest. The electoral roots of the system lie

in a first-past-the-post electoral rule, which established itself early on in England and the United States as the dominant mode of electoral representation. The generally acknowledged exemplar is the British Westminster system, as discussed.

Early critics of this system objected to the localist tendencies of the British electoral system, centered as it was on small (one- to two-member) constituencies. A proper political system, they thought, should act in the general interest, not in the interests of particular constituencies. PR reformers such as Leonard Courtney, Thomas Hare, Sir John Lubbock, and John Stuart Mill in England, Victor d'Hondt in Belgium, Carl Andrae in Denmark, Eduard Hagenbach-Bischoff in Switzerland, and Victor Considerant and A. Sainte-Lague in France were also bothered by the vulnerability of such a political system to the vagaries of popular opinion (Carstairs 1980; Noiret 1990). Because elections in a Westminster system rested on the votes of a few electors in swing districts party leaders had to test the current of public opinion carefully before taking the initiative. This led, it was charged, to a populist style of leadership, one oriented more toward pleasing the electorate than advancing its long-run interests (Hart 1992; Mill 1865/1958). Third, and most important, PR reformers objected to a system of election that effectively represented only two groups in parliament, and only one group in government. "In a really equal democracy," wrote J. S. Mill (1865/1958: 103–4), "every . . . section would be represented, not disproportionately, but proportionately . . . Man for man [the minority] would be as fully represented as the majority. Unless they are, there is not equal government, but a government of inequality and privilege: one part of the people rule over the rest."

The theory of centripetalism combines elements of the RPG model and criticisms leveled by PR reformers. The key to good governance, we propose, is not monopolization of power at the center but rather a flow of power from diverse sources toward the center, where power is exercised collectively. Two desiderata must be reconciled in order for this process of gathering-together to result in successful policies and policy outcomes. Institutions must be *inclusive*—they must reach out to all interests, ideas, and identities (at least insofar as they are relevant to the issue at hand). And they must be *authoritative*—they must provide an effective mechanism for reaching agreement and implementing that agreement. The concept of centripetalism thus implies both (a) broad-based inclusion and (b) centralized authority.

This is a problematic claim on the face of it. These two principles seem so radically opposed to each other that it is difficult to envision how a single institution, or set of institutions, could satisfy one criterion without sacrificing the other. They evoke dichotomies—masses versus elites, the people versus the state, small government versus big government, democracy versus autocracy, and, of course, Rousseau versus Hobbes. Granted, if governance is conceptualized in the usual way, as an arena in which interests are fixed and politics a zero-sum competition, then the notion of reconciling

inclusion and authority is polyannish. It seems fanciful to suggest that an institution could empower leaders without disempowering citizens.

We suppose, however, that interests are often constructed (endogenous), rather than primordial (exogenous). To be sure, the causal pathway between interests and institutions runs in both directions. But in the case of long-standing constitutional institutions such as those addressed here, the contemporary causal pathways are more likely to run from institutions to interests than vice-versa (Steinmo, Thelen, and Longstreth, 1992). Constitutional institutions condition the creation and reproduction of interests and identities. In particular, we expect that decentralist institutions establish a frame of reference in which identities and interests are conceptualized within a state/society dichotomy. Citizens are primed to see the state as a threat and civil society as an arena of liberty. Power is thus conceptualized in zero-sum terms: a stronger state means a weaker citizenry, a debilitated local community, or a "coopted" interest group. Centripetal institutions, by contrast, foster a positive-sum view of political power. Government is viewed as *creating* power, enhancing the ability of a political community through its chosen representatives to deliberate, reach decisions, and implement those decisions. Indeed, the authority of the centripetal state derives from its ability to bring together diverse groups and diverse perspectives under conditions of voluntary choice to a common meeting-ground, thus institutionalizing political conflict. Its power is persuasive, not coercive. Rather than a compromise position between inclusion and authority, we suggest that centripetal institutions actually reconcile these two principles, drawing the diverse strands of society together toward a single locus of sovereignty. The people rule, but they do so indirectly, through chosen representatives, and in a fashion that enhances rather than detracts from the authority of the state.

Centripetal institutions gather broadly; their roots are deep, that is, embedded. Through these institutions, diverse interests, ideas, and identities ("interests" for short) are aggregated. Particularistic interests are converted into ideologies; ideologies are converted into general-interest appeals; parochial perspectives are nationalized. Centripetal institutions thus encourage a search for common ground and culminate in an authoritative decision-making process, one not easily waylaid by minority objections. Institutions pull toward the center, offering incentives to participate and disincentives to defect. *Voice, not vetoes* is the motto of the centripetal theory of governance.

Visually, we may imagine the centripetal polity in a pyramidal shape—broad at the bottom and narrow at the top, with myriad connecting routes leading up, down, and across. Centripetal institutions thus establish an interlocked set of representative bodies stretching from the electorate at the base to the cabinet and prime minister at the apex. The electorate is represented in a legislature, which is in turn divided into committees, subcommittees, party caucuses, a cabinet, and perhaps various cabinet committees and commissions. At each stage of this process, a delegation of

power—a representational act—occurs. Tying each of these horizontal levels together is the vertical structure of the political party, the paradigmatic linkage mechanism.

This pyramidal structure fulfills the mandate of centripetalism—it gathers widely at the base, channeling interests, ideas, and identities upward to a single, authoritative policymaking venue. At each level, some narrowing of perspectives necessarily occurs. However, the pyramid encompasses a diversity of political parties as well as a variety of informal channels of communication. Through these channels—for example, special commissions, corporatist-style consultations, constituent-MP communications, hearings, ombudspersons, and so forth—nonpartisan messages can be heard (i.e., interests, ideas, and identities that do not fit neatly into the parties' missions). The centripetal polity thus "pulls" vertically and horizontally.

What, then, are the specific institutional features of the centripetal polity? The twin desiderata of inclusion and authority point to four constitutional-level features: unitary (rather than federal) sovereignty, unicameralism or weak bicameralism (i.e., a bicameral system with asymmetrical powers or congruent representation between the two houses), parliamentarism (rather than presidentialism), and a party-list proportional electoral system (rather than single-member districts or preferential vote systems). In addition, the centripetal polity should be characterized by a strong cabinet, medium-strength legislative committees, strong party cohesion, the power to dissolve parliament (no fixed terms), no limits on tenure in office, few elective offices, congruent election cycles, closed procedures of candidate selection (limited to party members), voting decisions largely dependent on the party identification of the candidate, party-centered political campaigns, multiparty (rather than two-party) competition, centralized and well-bounded party organizations, centralized and party-aligned interest groups, popular referenda only at the instigation of the legislature (or not at all), a restrained (nonactivist) judiciary, and a neutral and relatively centralized bureaucracy. Each of these institutional features serves to maximize, and if possible to reconcile, the twin goals of inclusion and authority, thus focusing power toward the center and gathering together diverse elements into a single policy stream.

Institutional contrasts with the decentralist model are summarized in Table 1. Note that the two models are different along all 21 dimensions. Sometimes the contrast is a matter of degrees and sometimes it is categorical. In any case, it is clear that we are faced with two opposing views of how to achieve good governance within a democratic framework.

Although one hesitates to rest any general theory on the status of individual countries, it may be heuristically useful to observe that although the United States is the generally acknowledged avatar of decentralism, and the United Kingdom the avatar of centralism, Scandinavia offers perhaps the best exemplars of centripetalism among the world's long-standing democracies. Sweden, Norway, and Denmark are all centripetal



**TABLE 1. Paradigms of Governance, Elaborated and Contrasted**

	Decentralism	Centripetalism
Territorial Sovereignty	Federal	Unitary
Legislative Branch	Bicameral, symmetrical, and incongruent	Unicameral, asymmetrical, or congruent
Executive	Presidential	Parliamentary
Electoral System	Single-member district or preferential vote	Party-list PR
Constitution	Written, with explicit limits on sovereignty	Unwritten or ambiguous, no explicit limits on sov.
Cabinet	Weak, durable	Strong, slightly less durable
Committees	Strong	Medium-strength
Party cohesion	Weak	Strong
Dissolution	No (fixed terms)	Yes
Term Limits	Perhaps	No
Elective Offices	Many	Few
Election Cycles	Incongruent	Congruent
Candidate Selection	Open, diffuse	Closed
Voting Cues	Personal vote	Party vote
Campaigns	Media, interest groups, candidate organization	Parties and party leaders
Party System	Two-party dominant	Multiparty
Party Organization	Weak, decentralized, porous	Strong, centralized, bounded
Interest Groups	Fragmented, nonpartisan	Centralized, party-aligned
Referenda	Possibly	No (or only at instigation of leg.)
Judiciary	Activist, independent	Restrained, Independent
Bureaucracy	Multiple independent agencies	Strong, neutral, relatively centralized

polities, as are a number of new or recently reformed democracies in Europe. Thus, the identification of centripetalism with the pattern of politics normal to continental Europe is an appropriate theoretical and empirical point of departure. However, there is no reason to limit the purview of this study to the OECD. Indeed, the rest of the democratic world, which now vastly outnumbers the OECD democracies, offers essential fodder for any empirical investigation that purports to be general in application.

## EMPIRICAL TESTS

It is not possible, nor would it be fruitful, to explore all 21 dimensions of centripetalism listed in Table 1. Of primary interest are those components of the centripetal theory that are measurable, exogenous (relative to other political institutions), and of presumed centrality to politics and policymaking. We refer to these factors as *constitutional*. They include the first four dimensions listed at the top of Table 1, demarcated by a dotted line: territorial sovereignty, the legislative branch, the executive, and the electoral system. Because the first two factors are closely related, both theoretically and empirically, we reduce this set to three: *unitarism*, *parliamentarism*, and *list-PR*.

We conceptualize unitarism along two dimensions: (a) the degree of separation (independence) between national and territorial units, and, if any separation at all, (b) the relative power of the two players (the more power the center possesses, the more unitary the system). Of the many institutional factors that determine variation along these dimensions, two predominate: federalism and bicameralism. A fully unitary polity should be both nonfederal and nonbicameral. Because these are matters of degree, however, we adopt a three-

part coding scheme for each dimension. *Nonfederalism* is coded as 0 = federal (elective regional legislatures plus constitutional recognition of subnational authority), 1 = semifederal (where there are elective legislatures at the regional level but in which constitutional sovereignty is reserved to the national government), or 2 = nonfederal. *Nonbicameralism* is coded as 0 = strong bicameral (upper house has some effective veto power; the two houses are incongruent), 1 = weak bicameral (upper house has some effective veto power, though not necessarily a formal veto; the two houses are congruent), or 2 = unicameral (no upper house or weak upper house). The unitarism variable is constructed by averaging the scores of these two components together.<sup>2</sup>

*Parliamentarism* is understood as a system of government in which the executive is chosen by, and responsible to, an elective body (the legislature), thus creating a single locus of sovereignty at the national level. *Presidentialism*, its contrary, is a system where policymaking power is divided between two separately elected bodies, the legislature and the president. The president's selection is usually by direct popular election, though it may be filtered through an electoral college (as in the United States), and the rules pertaining to victory (i.e., by relative or absolute majority) vary from country to country. His or her tenure cannot be foreshortened by parliament except in cases of gross malfeasance. She or he is actively engaged in

<sup>2</sup> The combination of these two dimensions is justified by the fact that they are linked empirically (constitutional federalism is a necessary condition for strong bicameralism) and conceptually (the purpose of a strong second chamber is usually to protect the powers and prerogatives of subnational units). In a fully unitary state, territorial units (if any) have no constitutional standing, no independently elected territorial legislature, no specific policy purviews reserved to them, and minimal revenue-raising authority.

the making of public policy, and in this sense plays a political (i.e., partisan) role. In practice, between these two polar types we find many admixtures, known generically as semipresidential systems. Thus, we conceptualize the parliamentary/presidential distinction as a continuum with two dimensions: (a) the *degree of separation* (independence) between president and parliament (unity = parliamentary, separation = presidential), and, if there is any separation at all, (b) the *relative power* of the two players (the more power the president possesses, the more presidential is the resulting system). We capture this complex reality with a three-part coding scheme: 0 = presidential, 1 = semipresidential, 2 = parliamentary.

The centripetal theory of democratic governance suggests that electoral systems, like other constitutional elements of a polity, should maximize the twin desiderata of authority and inclusion. These twin goals are best achieved when an electoral system encourages strong national parties while also maintaining low barriers to entry for new parties, strong competition among existing parties, and demographically diverse party delegations. This, in turn, mandates an electoral system that privileges interparty choice and intraparty representation over intraparty electoral choice. Voters vote, and parties nominate. Further, the vote choice itself should be based on national, partisan principles rather than on preferences for individual candidates or district-level concerns. Insofar as “personality” matters, it should be the personality of the party leader, not the district-level candidate, that influences voter choices. Empirically, three features of an electoral system bear critically on these issues: (a) district magnitude (*M*), (b) seat allocation rules (majoritarian or proportional), and (c) candidate selection rules. The centripetal ideal type is defined by  $M > 1$ , proportional seat allocation rules, and party-controlled candidate selection. This is the familiar *closed-list-PR* electoral system—“list-PR” for short. Other systems are ranked lower in this coding according to their deviation from this ideal type. Thus, the coding for the list-PR variable is as follows: 0 = majoritarian or preferential-vote, 1 = mixed-member majority (MMM) or block vote, and 2 = closed-list PR.

Granted, it takes time for institutions to exert an appreciable effect on governance outcomes. A country switching from a presidential system to a parliamentary system (or establishing a parliamentary system in a newly democratic or independent setting) should not expect to see immediate, dramatic changes in the quality of governance. Instead, these effects are likely to cumulate over time as new institutional rules begin to condition actions and expectations. History matters, though recent history should matter more. To this end, we create a moving, weighted sum of each country's annual unitarism, parliamentarism, and list-PR scores, beginning in 1901 and ending in the observation year. The weights are constructed so as to capture long-term historical patterns while giving greater weight to more recent years. A country's weighted-sum unitarism (or parliamentarism or list-PR) score in 1980 is the weighted sum of its scores from 1901 to 1980. Its score

in 1981 is the weighted sum of its scores from 1901 to 1981, and so on.<sup>3</sup>

Because our theoretical interest is in the combined effect of unitarism, parliamentarism, and list-PR, we create a final composite variable, *Centripetalism*, by adding together the historical, weighted-sum scores for these three variables (equally weighted). Henceforth, when referring to the variable, we capitalize this term (Centripetalism); when referring to the theory or concept of centripetalism, we do not.

Country-years figure in this coding process, and in the empirical analyses to follow, so long as a country surpasses a minimum threshold of democracy during a given year. Recall that centripetalism is a theory of *democratic* governance; it has no application within authoritarian settings. We employ a relatively low threshold of democracy because we wish to include as many plausible cases as possible in our analysis and because we expect the logic of centripetalism to be operative so long as there is a modicum of multiparty competition. A country-year counts in our empirical analysis and in the weighted summation process so long as it obtains a score greater than zero (on a scale ranging from -10 to 10) on the Polity2 democracy indicator.<sup>4</sup>

<sup>3</sup> The weights used change progressively by the observation year used in the analysis. For the observation year 1980, for example, a country's raw score in 1901 is weighted by 1/80, its score in 1902 by 2/80, its score in 1903 by 3/80, until finally reaching a weight of 1 (80/80) in 1980. Each of these weighted, annual scores is summed for a single country into a cumulative score for a given observation year. (For the observation year 1981, the weighting denominator would be 81, and so on.) The formula for the weighting schemes is as follows: Let *S* be the raw score, and *W* the weighted score, then

$$W_t = \sum_{s=1901}^t \left( \frac{s-1900}{t-1900} \right) * S_s.$$

For the observation years 1980 and 2000, for example, the weighting schemes would be.

$$\begin{aligned} W_{1980} &= \sum_{s=1901}^{1980} \left( \frac{s-1900}{80} \right) * S_s = \frac{1}{80} S_{1901} + \frac{2}{80} S_{1902} \\ &\quad + \dots + \frac{79}{80} S_{1979} + \frac{80}{80} S_{1980} \\ W_{2000} &= \sum_{s=1901}^{2000} \left( \frac{s-1900}{100} \right) * S_s = \frac{1}{100} S_{1901} + \frac{2}{100} S_{1902} \\ &\quad + \dots + \frac{99}{100} S_{1999} + \frac{100}{100} S_{2000}. \end{aligned}$$

If a country is nondemocratic (receiving a Polity2 score of less than 0) in a given year, or if a country is not formally sovereign during that year, it receives a score of 0 for that year.

<sup>4</sup> Marshall and Jaggers (2005). Because the Polity2 democracy score does not contain data for several countries (mostly micro-states), we impute missing values using the following alternative measures of democracy: the Freedom House Political Rights indicator (Piano and Puddington 2004), Bollen's (1993) Liberal Democracy variable, Vanhanen's (1990) Competition measure, and Banks's (1994) Legislative Effectiveness I and II and Party Legitimacy variables. A complete list of country cases that meet our minimal definition of democracy, along with their weighted, historical Centripetalism scores, and their annual (raw) scores on all four component variables

## Operationalizing Good Governance

A general model of good governance applies (by definition) to any outcome that may be deemed, on the whole, good or bad for a society, that is, for or against the public interest. There is plenty of room for debate on these matters. Even so, we suppose that consensus can be reached on the normative valence of many policies and policy outcomes. It is on this tentative consensus—itsself contingent on evidence and further normative reflection—that any empirical study of good governance rests.<sup>5</sup>

In this paper, we limit ourselves to a consideration of three broad policy areas: political development, economic development, and human development. Within these three areas we explore eight specific measures of good/bad governance: (1) bureaucratic quality, (2) tax revenue, (3) investment rating, (4) trade openness, (5) gross domestic product (GDP) per capita, (6) infant mortality, (7) life expectancy, and (8) illiteracy. We choose these indicators over others because they offer evidence of broad patterns of governance and because they allow for longitudinal analysis across several decades and latitudinal analysis across most of the democracies in the world. They are ideal, in other words, for time-series cross-section analysis. We do not suppose that they exhaust the field of governance indicators, merely that they offer a useful collection of indicators of valued outcomes across an array of policy areas.

*Bureaucratic quality*, a measure of political development, is an indicator ranging from 0 to 6 (with higher scores indicating higher quality) developed by the Political Risk Services (PRS) group as part of its International Country Risk Guide (ICRG). It gauges the institutional strength and quality of the civil service, measured along six dimensions: adequate pay, independence from political pressures, professionalism (adequate training, recruitment by merit rather than by patronage), capacity (ability to respond to assigned tasks), appropriate staffing (neither over- nor

understaffed), and freedom from corruption (Howell 1998, 194).

*Tax revenue* is a “hard” (objectively quantifiable) measure of political development. A government’s capacity to extract resources from businesses and individuals should reflect its overall capacity to formulate and implement public policies (Cheibub 1998; Lieberman 2002). (We control for natural resource wealth in subsequent tests, so tax revenue does not reflect the existence of “easy money” in the form of oil or diamond receipts.) The variable employed here, drawn from the World Bank’s *World Development Indicators* (World Bank 2003), measures aggregate tax revenues, considered as a share of GDP. More specifically, it counts compulsory, unrequited, nonrepayable receipts for public purposes collected by the central government, including interest collected on tax arrears and penalties collected on nonpayment or late payments of taxes.

*Investment rating* measures the safety to potential investors of acquiring a stake in a country’s economy. Many academics regard it as a proxy for the quality of economic policy in a country; the higher the rating, the lower the risk and the better a government’s economic policies are thought to be. In recent years, risk assessment has become a substantial business, a sideline for most consulting firms with international clients. Consequently, there are a variety of investment rating indicators to choose from. Among these, *Euro-money’s* country risk index enjoys perhaps the most comprehensive coverage. (Reassuringly, it correlates strongly with other indices.) *Euromoney* ratings are based on polls of economists and political analysts and supplemented by quantitative data such as debt ratios and access to capital markets. The overall country rating derives from nine separate categories, each with an assigned weighting (in parentheses): (1) political risk (25%); (2) economic performance (25%); (3) debt indicators (10%); (4) debt in default or rescheduled (10%); (5) credit ratings (10%); (6) access to bank finance (5%); (7) access to short-term finance (5%); (8) access to capital markets (5%); and (9) discount on forfeiting (5%) (Euromoney 2004).

*Trade openness* is measured by the sum of total imports and exports, expressed as a share of GDP (logarithm, data source: World Bank 2003). This indicator reflects, in part, the degree to which a country opens its borders to trade; it is thus a policy measure, not simply a policy outcome. Indeed, a host of poor and inefficient economic policies, including high tariff and nontariff barriers, poorly managed exchange rates, and corruption in the customs bureau are likely to depress the growth of imports and exports. Although there is debate over the relative impact of trade on overall economic growth performance (see Krueger 1995; Rodrik 1995), few economists would argue that trade depresses growth rates. In addition, we have found in our own work that trade may have positive effects on human development, even when controlling for economic performance (Gerring, Thacker, and Moreno N.d., c). Thus, there are strong grounds for regarding trade as an indicator of good governance.

(unitarism, parliamentarism, list-PR, and Centripetalism), can be found at <http://www.bu.edu/sthacker/data.html>. For a more detailed explanation of coding procedures, see Gerring and Thacker N.d.

<sup>5</sup> One may contrast the approach taken here with alternative approaches to governance focused on several closely related concepts: “public goods,” “rent,” “pareto optimality,” and “efficiency.” We find these approaches to be highly ambiguous, hence, resistant to operationalization. Moreover, although these approaches purport to be value-neutral, they often smuggle in some conception about what is, and is not, in the public interest. In this respect, the biggest difference between such approaches and our own is the degree of normative transparency. Finally, we argue that on those occasions where these concepts do not conform to common notions about the public interest, they are, by definition, not useful as policymaking guides. Thus, either (a) the notion of a public good is equivalent to the notion of a policy that advances the public interest, in which case our approach is equivalent to the conventional approach; or (b) these two notions diverge, and the concept of a public good becomes ambiguous, not to mention tenuous (how can a policy provide public goods and not also advance the public interest?) In sum, we believe that an explicitly normative theory of governance is not only possible but also unavoidable if political science is to be of any use to policymakers (see Gerring and Thacker N.d.).

*GDP per capita* is a measure of average income levels, or the real value of total production within an economy during the course of a year divided by the total population (logarithm, data source: World Bank 2003). We measure economic performance as a level, rather than as a change (i.e., growth) variable because our interest lies in the level of prosperity attained in a given country, rather than in its short-run rate of change. This is also in keeping with our approach to other governance outcomes; for example, we measure the level of trade from year to year, not the change in trade from year to year.

*Infant mortality* is measured by the infant mortality rate (IMR), the number of deaths per one thousand lives births that occur in the first year of life (logarithm, data source: World Bank 2003). IMR, a primary measure of human development, is affected by many government policies (particularly social policies) and is thus an important outcome-based measure of good governance.

*Life expectancy* measures the expected tenure of life in a country at birth, extrapolating from mortality statistics available at that time (Bos, Vu, and Stephens 1992; Riley 2001; logarithm, data source: World Bank 2003). Like IMR, life expectancy is an overall measure of human development strongly influenced by government policies; hence, it provides a good indicator of the quality of governance in a country.

*Illiteracy* is measured as the percentage of people age 15 and older who cannot, with understanding, both read and write a short, simple statement on their everyday life (logarithm, data source: World Bank 2003). Literacy has become a standard feature of human development indices in recent decades and largely reflects the success of government-sponsored education policies.<sup>6</sup>

## Research Design

Because the theory of centripetalism is applicable only within a democratic framework, we limit all regression analyses to country-years that are minimally democratic, as discussed earlier. Resulting samples vary from a minimum of 77 countries to a maximum of 126, and from a minimum of 14 years to a maximum of 4 decades (1960–2000).

The literature on the various topics captured in our eight dependent variables suggests the inclusion of 15 core controls in the following analyses. We include a

time trend variable to control for spurious correlation between any pair of similarly trended dependent and independent variables; this should be signed in whatever direction a given dependent variable is trended, on average, over time. To capture a country's regime history, we employ a variable that measures democracy stock historically. We construct this variable by taking the logarithm of the sum of a country's Polity2 scores (Marshall and Jaggers 2005) from 1900 to the observation year (for further discussion see Gerring, Bond, and Barndt N.d). We anticipate this variable to have a positive association with good governance. GDP per capita (logarithm, World Bank 2003) should also be associated with better governance outcomes. Dummy variables for Africa and Latin America/Caribbean are expected to reflect lower levels of governance in those regions compared to others, whereas expectations for Asia are mixed (e.g., better bureaucratic quality, but lower tax revenues). We anticipate that a significant period of socialist rule (LaPorta et al. 1999) has negative effects on bureaucratic quality, investment rating, trade openness, and GDP per capita, and positive effects on the remaining governance indicators. Having an English legal origin is often thought to promote good governance (LaPorta et al.). To the extent that countries farther from the equator have better governance, latitude (absolute value, scaled to 0–1, logarithm, LaPorta et al.) should correlate with better outcomes. Expectations for ethnic (and linguistic) fractionalization (Alesina et al. 2002) are more tentative; however, heterogeneity is generally expected to hamper the quality of governance in a country. To the extent that having a large population (total population, logarithm, World Bank 2003) makes certain governmental tasks more difficult, population might be expected to diminish governance quality. Distance (in thousands of kilometers) from the nearest financial center (Tokyo, New York, or London) is intended to capture the negative impact of geographic distance from the "cores" of the international economy. Oil (millions of barrels per day per capita) and diamond (rescaled to billions of metric carats per year per capita) production levels capture the "resource curse" (Humphreys 2005). Yet, these resources also provide sources of revenue and wealth. As such, expectations are mixed.<sup>7</sup>

We also include a control variable that measures the average value of the dependent variable across all countries, weighted by the inverse of the geographic distance (in kilometers) of each country from the country in question. (In the case of GDP per capita, we weight the average value of the dependent variable by each country's share of trade with the observed country, rather than by the inverse of the geographic distance between the countries.) Countries lying close to one another may display similar values for extraneous reasons

<sup>6</sup> Descriptive statistics for the Centripetalism variable, all dependent variables, and the control variables can be found at <http://www.bu.edu/sthacker/data.html>, along with a correlation matrix for all variables. Note that missing data for variables measuring infant mortality and life expectancy are interpolated so as to reduce sample bias. Because these variables are all heavily trended, we do not anticipate that interpolation introduces new systematic biases in the data. For illiteracy, missing data from the WDI dataset—primarily for the OECD countries—are imputed using Banks 1994. Some missing data are also extrapolated, with the assumption that once a level of 0.01% illiteracy is attained it remains constant through time. For GDP per capita, small amounts of missing data are imputed using Penn World Tables 6.1 (Heston, Summers, and Aten 2002).

<sup>7</sup> Some indicators measure the export value of these last two items as a percentage of all exports or of GDP. We believe that this confuses two issues—the extent of natural resources in a country and the degree of its economic development or export orientation, which is implicit in the denominator. Because it is the first, not the second, that we wish to measure we employ a "raw" measure of natural resources

(culture, geography, diffusion, and so forth). Thus, we anticipate a positive sign for this variable. The inclusion of this variable in all regressions should help minimize possible spatial autocorrelation in the sample.

We employ additional variables in selective regressions, as appropriate. We include Protestants in the analysis of bureaucratic quality and Muslims in the estimations for various human development outcomes (both are measured as a percent of the total population). Prior research suggests that a Protestant heritage may improve state capacity (Gerring and Thacker 2004), whereas having a large Muslim population may impede human development (Moon 1991). Linguistic fractionalization (Alesina et al. 2002) substitutes for ethnic fractionalization in our analysis of illiteracy, for obvious reasons.

Because there exists no standard benchmark model for any of these regressions we conduct two tests for each dependent variable. The first is a full model, including all variables discussed above. The second is a reduced-form model that sequentially deletes variables that do not pass a minimal threshold of statistical significance ( $p < 0.10$  in two-tailed tests), in the expected direction. We retain several controls (the geography-weighted dependent variable, the time trend, democracy stock, and GDP per capita) in all models, regardless of statistical significance, because of our strong prior expectation that these variables capture important and otherwise unobserved effects. Reassuringly, the key variable, Centripetalism, remains quite stable across both full and reduced-form specifications.

To minimize possible endogeneity between left- and right-side variables (and among certain right side variables), we measure two indicators in the first year of our sample (1960), rather than on an annual basis. This applies to GDP per capita and population. Where we are less concerned about endogeneity we allow indicators to vary from year to year, but lag them by 1 year (except in the case of the geography-weighted dependent variable, which is contemporaneous). We treat other controls, such as region, socialism, legal origin, fractionalization, distance from the nearest financial center, Protestantism and Muslim, as constant through time.

Time-series cross-section (TSCS) regression calls forth two broad specification issues (among others). Datasets that employ both cross-national and time-series data are subject to potentially stubborn problems, spatially and temporally. Regrettably, we cannot employ a unit-based fixed-effect research design to address spatial issues, such as unobserved heterogeneity, because our causal variable, Centripetalism, does not vary sufficiently from year to year (Beck 2001, 285; Beck and Katz 2001, 492–93). We do, however, employ a set of regional “fixed effects” and a geographically weighted version of the dependent variable (see earlier) to help remedy spatial problems. With respect to temporal issues, we employ a statistical correction for first-order autocorrelation and a time-trend variable to control for possibly spurious correlations between a heavily trended dependent variable and a similarly trended independent variable.

All regressions employ Newey–West standard errors, which assume a heteroskedastic error distribution and apply a TSCS equivalent of Huber/White/sandwich, or “robust,” standard errors, along with a correction for first-order autocorrelation (Newey and West 1987). Although Newey–West is a common approach in economics, it is less frequently used in political science. We employ it here because it achieves the aforementioned goals and is somewhat less computationally expensive than the alternatives. In any event, results are equivalent in other formats, for example, with a Prais–Winsten feasible generalized least squares (FGLS) approach with panel corrected standard errors and an AR1 correction for autocorrelation.

## Results

Table 2 shows estimation results for each of the eight governance indicators. The fit of the models is strong in each case, with F-values significant at better than 0.0001, and a pseudo- $R^2$  that ranges from 0.58 to 0.90.<sup>8</sup> Collectively, these models are highly significant and our predicted values fit the actual values well. Control variables generally behave as expected, though they are not always statistically significant. The geographically weighted dependent variable control is correctly signed and significant in most, but not all, specifications. Results for the time trend variable suggest that our measures of governance tend to improve over time on average, and most are statistically significant. Similarly, countries with a long democratic history (or stock) show better patterns of governance, with significant results in most cases. Findings for the GDP per capita baseline (1960) measure confirm that countries that start out wealthy tend to end up with higher quality governance. The results for the regional dummies are usually consistent with theoretical expectations. A period of socialist rule is generally associated with improved human development, poor economic performance, and high tax revenues. An English legal origin is associated with good governance across several indicators. Findings for fractionalization, baseline (1960) population size, and distance from the nearest financial center confirm expectations in most specifications, whereas those for latitude and oil and diamond production are less consistent. Protestantism is associated with more bureaucratic quality, whereas having a large Muslim population seems to hamper human development.

With respect to our theory, we find strong support for the centripetal model of governance. The coefficient for Centripetalism is correctly signed and

<sup>8</sup>  $R^2$  is an ordinary-least-squares concept and we are using a generalized least-squares estimator. In order to report a measure of fit, we calculate a pseudo- $R^2$  equal to the square of the correlation between the dependent variable and its predicted values in each equation. Note that the use of the geography-weighted and time trend control variables likely inflates the pseudo- $R^2$  values obtained here. We report them as a measure of fit for the interested reader, without placing much substantive emphasis on them.

**TABLE 2. Empirical Tests**

	Bureaucratic Quality (+ = good gov.)		Tax Revenue (+ = good gov.)		Investment Rating (+ = good gov.)		Trade Openness (ln) (+ = good gov.)		GDPpc (ln) (+ = good gov.)	
	1	2	3	4	5	6	7	8	9	10
Centripetalism	<b>0.002***</b> (0.001)	<b>0.001</b> (0.001)	<b>0.036***</b> (0.004)	<b>0.036***</b> (0.004)	<b>0.017**</b> (0.008)	<b>0.017**</b> (0.008)	<b>0.001***</b> (0.0003)	<b>0.002***</b> (0.0003)	<b>0.001**</b> (0.0002)	<b>0.001**</b> (0.0002)
Geo-weighted control	-0.001 (0.009)	0.002 (0.008)	0.079*** (0.009)	0.074*** (0.007)	0.005 (0.005)	0.005 (0.004)	0.0002 (0.001)	0.002* (0.001)	0.040*** (0.015)	0.039*** (0.014)
Trend	0.017* (0.010)	0.014 (0.010)	0.006 (0.024)	0.012 (0.023)	0.327*** (0.086)	0.314*** (0.085)	0.008*** (0.001)	0.008*** (0.001)	0.015*** (0.001)	0.015*** (0.001)
Democracy stock (ln)	0.354* (0.197)	0.466** (0.185)	-0.539 (0.909)	-0.096 (0.879)	14.588*** (2.393)	15.041*** (2.147)	-0.062 (0.051)	-0.220*** (0.046)	0.448*** (0.112)	0.433*** (0.091)
GDP per cap (1960)	0.825*** (0.051)	0.754*** (0.042)	1.439*** (0.290)	1.111*** (0.230)	9.005*** (0.579)	8.943*** (0.552)	-0.009 (0.014)	-0.027*** (0.010)	0.786*** (0.064)	0.771*** (0.052)
Africa	0.040 (0.196)		1.679 (1.059)		-4.294** (1.701)	-5.092*** (1.550)	-0.136*** (0.049)		-0.193** (0.098)	-0.274*** (0.068)
Asia	0.635*** (0.182)	0.842*** (0.148)	-3.158*** (0.698)	-3.659*** (0.649)	8.109*** (1.808)	8.286*** (1.842)	0.065 (0.050)		0.145 (0.095)	
Latin Am/ Carib	-0.876*** (0.153)	-0.938*** (0.127)	-1.457** (0.581)	-1.856*** (0.528)	-12.368*** (1.421)	-12.650*** (1.374)	-0.271*** (0.033)	-0.216*** (0.023)	0.072 (0.050)	
Socialism	0.389* (0.204)		5.342*** (1.002)	5.054*** (1.005)	-9.645*** (2.215)	-9.776*** (2.232)	0.296*** (0.051)		-0.277*** (0.099)	-0.316*** (0.105)
English legal origin	0.564*** (0.098)	0.501*** (0.102)	2.879*** (0.479)	2.788*** (0.485)	0.661 (1.031)		0.072*** (0.026)	0.102*** (0.025)	-0.012 (0.036)	
Latitude (ln)	-0.159* (0.084)		-0.283 (0.373)		2.265*** (0.851)	2.450*** (0.759)	-0.042** (0.019)		0.215*** (0.045)	0.203*** (0.044)
Ethnic fract.	0.408** (0.185)		-5.604*** (1.087)	-4.516*** (0.917)	-1.972 (2.024)		0.394*** (0.049)		-0.275*** (0.074)	-0.274*** (0.077)
Population (ln) (1960)	0.138*** (0.032)		-0.484*** (0.102)	-0.583*** (0.108)	2.869*** (0.248)	2.762*** (0.225)	-0.242*** (0.007)	-0.216*** (0.007)	-0.021*** (0.008)	-0.016** (0.007)
Distance fin center	-0.026 (0.017)		0.064 (0.105)		-0.926*** (0.232)	-0.902*** (0.226)	-0.043*** (0.005)	-0.046*** (0.004)	-0.048*** (0.005)	-0.046*** (0.006)
Oil prod.	-1.957*** (0.562)	-1.771*** (0.479)	-0.386 (2.347)		-0.679 (2.566)		-0.002 (0.096)		0.393** (0.160)	0.410** (0.163)
Diamond prod.	0.130*** (0.021)		0.007 (0.155)		1.496*** (0.212)	1.536*** (0.208)	0.019*** (0.006)		0.133*** (0.007)	0.132*** (0.007)
Protestant (%)	0.007*** (0.002)	0.006*** (0.001)								
Constant	-8.302*** (1.379)	-5.876*** (1.146)	14.602** (5.785)	16.280*** (5.813)	-160.214*** (15.129)	-160.958*** (14.551)	7.969*** (0.361)	8.998*** (0.316)	-0.825* (0.451)	-0.657 (0.413)
Observations	716	733	1643	1663	1544	1576	2521	2609	2522	2522
Countries	77	79	105	106	122	124	126	131	124	124
Sample Period	1981-94	1981-94	1969-00	1969-00	1981-00	1981-00	1960-00	1960-00	1960-09	1960-99
Pseudo R <sup>2</sup>	0.80	0.78	0.58	0.58	0.80	0.81	0.68	0.63	0.90	0.90
p > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

significant at the 90% level of confidence or better in all estimations but one. Reassuringly, these results are robust to the inclusion of a wide variety of control variables, as evidenced by the similarities in coefficients and standard errors for Centripetalism across full and reduced-form models. Centripetalism is associated with more bureaucratic quality, higher tax revenues, better investment ratings, more trade openness, greater economic prosperity, fewer infant deaths, longer life expectancy, and lower rates of illiteracy, all else being equal.

A number of other possible variables might have been employed as controls in this wide-ranging series of regression tests. Indeed, we tested a much larger

range of theoretically plausible control variables in the course of the analysis than could be included in the accompanying tables. These specification tests included income inequality (Gini coefficient), state history (the length of time a country has been independent), decade dummies (to further control for time effects), and additional measures of population heterogeneity and colonial history (for a complete list see Gerring and Thacker N.d.). Our central results were robust in each of these tests, and none revealed any systematic patterns that warranted inclusion of additional controls in our benchmark equation. We have confidence in the specifications presented in Table 2, not because we imagine to have discovered the one "true" equation for

TABLE 2. Continued

	IMR (ln) (- = good gov.)		Life Expectancy (ln) (+ = good gov.)		Illiteracy (ln) (- = good gov.)	
	11	12	13	14	15	16
Centripetalism	-0.001*** (0.0002)	-0.001*** (0.0002)	0.0001* (0.00004)	0.0001*** (0.00004)	-0.001* (0.0004)	-0.001** (0.0003)
Geo-weighted	0.002* (0.001)	0.003*** (0.001)	0.00002 (0.0001)	-0.0002* (0.0001)	0.056*** (0.008)	0.058*** (0.008)
control						
Trend	-0.031*** (0.001)	-0.031*** (0.001)	0.003*** (0.0002)	0.004*** (0.0002)	-0.020*** (0.002)	-0.020*** (0.002)
Democracy	-0.546*** (0.068)	-0.554*** (0.060)	0.032*** (0.011)	0.013** (0.007)	-0.534*** (0.126)	-0.474*** (0.124)
stock (ln)						
GDP per cap	-0.304*** (0.035)	-0.298*** (0.024)	0.047*** (0.006)	0.045*** (0.005)	-0.746*** (0.031)	-0.769*** (0.025)
(1960)						
Africa	0.395*** (0.070)	0.452*** (0.055)	-0.191*** (0.014)	-0.203*** (0.013)	0.211** (0.085)	0.255*** (0.067)
Asia	-0.117 (0.084)		0.026** (0.012)		-0.133 (0.096)	
Latin Am/	0.226*** (0.051)	0.293*** (0.041)	0.015* (0.008)		0.518*** (0.089)	0.596*** (0.066)
Carib						
Socialism	-0.453*** (0.078)	-0.462*** (0.080)	0.019 (0.012)		-2.044*** (0.157)	-2.016*** (0.149)
English legal	-0.020 (0.027)		-0.013** (0.006)		-0.329*** (0.052)	-0.331*** (0.051)
origin						
Latitude (ln)	-0.051 (0.034)		0.017*** (0.006)	0.015*** (0.005)	-0.073* (0.044)	
Ethnic fract.	0.482*** (0.061)	0.495*** (0.053)	-0.064*** (0.011)	-0.064*** (0.012)		
Population (ln)	0.028*** (0.008)	0.026*** (0.007)	-0.007*** (0.001)	-0.005*** (0.001)	0.017 (0.011)	
(1960)						
Distance finance	0.041*** (0.005)	0.040*** (0.005)	0.001 (0.001)		0.060*** (0.008)	0.062*** (0.008)
center						
Oil prod	-0.321*** (0.117)	-0.290*** (0.108)	0.002 (0.025)		-0.219 (0.275)	
Diamond	0.001 (0.010)		-0.003 (0.005)		0.012* (0.006)	
prod						
Muslim	0.005*** (0.001)	0.005*** (0.001)	-0.001*** (0.0001)	-0.001*** (0.0001)	0.009*** (0.001)	0.009*** (0.001)
(%)						
Linguistic fract.					0.703*** (0.091)	0.691*** (0.091)
Constant	9.285*** (0.389)	9.342*** (0.373)	3.702*** (0.063)	3.825*** (0.039)	10.504*** (0.872)	10.548*** (0.811)
Observations	2633	2663	2634	2652	2401	2438
Countries	126	127	125	127	108	110
Sample Period	1960-00	1960-00	1960-00	1960-00	1960-00	1960-00
R <sup>2</sup>	0.82	0.82	0.75	0.74	0.85	0.85
p > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Newey-West regression. Standard errors in parentheses. Sample limited to country/years that are minimally democratic (Polity2 > 0).

\*p < 10%; \*\*p < 5%; \*\*\*p < 1%.

each outcome, rather, because the inclusion (or exclusion) of a wide variety of plausible controls does not substantially alter the results with respect to our main hypothesis.

Of course, we realize that the "treatment" is not randomized. It is possible, for example, that centripetal institutions are more likely to be adopted where prospects for good governance are otherwise more propitious, in which case our key variable may be proxying for other, unmeasured factors. In order to gauge the robustness of our findings in the face of this identification problem, we employed a series

of instruments for Centripetalism in two-stage least-squares estimations of the same set of governance outcomes as shown in Table 2.<sup>9</sup> Results from these instrumental variables estimations are at least as strong as the findings presented in Table 2—and in some

<sup>9</sup> Chosen instruments for Centripetalism include democracy stock (logged), latitude, ethnic fractionalization, religious fractionalization (Alesina et al. 2002), Western Europe (dummy), state history, social conflict (a compilation of measures from Marshall 1999), instability (a compilation of measures from Banks 1994), and population size (1960, logged).



cases stronger—thus providing some assurance that the effects reported here are not simply the product of nonequivalent treatment and control groups. Yet, we do not have a great deal of confidence in the two-stage models. All of the possible instruments available to us violate at least one of the assumptions of instrumental-variable analysis (Reiss 2003): they are either poorly correlated with Centripetalism, or they are correlated with the error term (i.e., they are probable causes in their own right of good/bad governance in the contemporary period). Thus, although the instrumental variable results provide support for the validity of our central findings, we do not regard this technique as an appropriate one for the present analysis (hence, we do not report the results here).

In any case, we think it unlikely that the choice of constitutional institutions reflects a country's future prospects for good (bad) governance. To be sure, whether a country becomes unitary or federal, parliamentary or presidential, list-PR or majoritarian depends partly on a country's colonial heritage, its size and heterogeneity, and on distinctive regional or historical patterns. However, these exogenous influences are relatively easy to model and appear as controls in all our regression tests. Other factors influencing constitutional choice are more or less stochastic and do not seem to accord with a country's proclivity to good or bad governance. In some instances, for example, federal institutions have been chosen because of their anticipated success in resolving conflict among heterogeneous groups (e.g., Canada, India, Switzerland, the United States). In other instances, unitarism has been viewed as the cure for precisely the same set of conflicts. This is the approach taken by all currently unitary states, whose populations were once—and in many cases remain—fractious and diverse (e.g., France, United Kingdom). In short, it all depends.<sup>10</sup> It is not the case, therefore, that federalism is chosen only in instances of high conflict or great heterogeneity.

One must also consider the fact that constitution-makers generally have notoriously short time-horizons. They are usually interested in installing a system that will benefit them personally, their parties, or their constituencies. In this respect, the type of constitution a country arrives at is the product of a highly contingent political battle, with no bearing on a country's long-term governance potential. Finally, one must reckon with the dubious assumptions made by each contending group (or by voters, if the agreement is ratified by the populace). Presidential systems, for example, are commonly viewed as installing "strong" government; however, most political scientists believe that parliamentarism fosters energy and efficiency in the executive. Thus, even where calculations by constitution-makers

extend to the long-term quality of governance in a polity, they are of dubious significance in achieving that result. Precisely because framers do not know which constitutional factors lead to good governance, whatever wisdom and far-sightedness they may possess is of little practical import. For all these reasons, we think it fair to regard a country's choices among constitutional institutions as a largely stochastic phenomenon with respect to the outcomes of interest in this study: long-term patterns of good or bad governance.

We must stress, finally, that the arguments presented in this short article are neither final nor incontrovertible. Many additional measures of good governance might be probed. Many additional specification tests might be conducted. A higher threshold of democracy (e.g., Polity2 > 4) might be employed, producing a somewhat smaller sample of "high-quality" democracies. Additional statistical formats, correcting for a variety of spatial and temporal problems, might be applied. The components of Centripetalism—unitarism, parliamentarism, and list-PR—might be individually assessed. Different weighting schemes might be employed to capture the historical effect of these political institutions. We pursue these and other issues elsewhere (Gerring and Thacker N.d.; Gerring, Thacker, and Moreno N.d., a and N.d., b).

Even so, the results in this paper indicate that centripetal institutions are associated with good governance across a wide range of indicators of political, economic, and human development. Moreover, it seems plausible to suppose that these results are indicative of a probabilistic causal relationship between centripetal constitutions and good governance.

In evaluating the practical impact of centripetal institutions, a few examples drawn from a hypothetical scenario may be useful. Employing the coefficients for Centripetalism from the full-form models in Table 2 (and keeping all control variables constant), we find that 50 years of fully centripetal institutions (1951–2000, in this example) are associated with an improvement (compared to the fully decentralist polity) of 0.54 points on the 7-point scale of bureaucratic quality, more than 8% of GDP in tax revenues, nearly 4 points higher investment rating (on a 100-point scale), a 32% increase in trade, 13% higher GDP per capita, 23% fewer infant deaths, 1.5% longer life expectancy, and 15% lower rates of illiteracy.<sup>11</sup>

Although these figures are only illustrative, they suggest that the relationship between Centripetalism and good governance has substantive importance. To be sure, the causal effect of Centripetalism on any single outcome may be relatively modest, and it is by no means inconsequential. Relative to the other independent variables explored in Table 2, Centripetalism has the most consistent and, judging by standardized

<sup>10</sup> This raises another possibility. Perhaps unitarism is a sign of successful state building, rather than a cause. Yet, this flies in the face of many countries' experiences. Federal constitutions have proven successful in establishing strong nation-states in Switzerland and the United States, unitary constitutions have proven less successful in Spain, Italy, and the United Kingdom, as witnessed by recent trends toward greater devolution and persistent regional dis-harmony.

<sup>11</sup> The coefficients for the logged dependent variables (trade, GDP per capita, infant mortality, life expectancy, and illiteracy) reported in Table 2 measure the effect of a 1-unit change in the independent variable on those outcomes as a percentage change in the dependent variable. Thus, a 1-unit change in the independent variable results in a change in the dependent variable of 100 \*  $\beta$ % (Wooldridge 2002).

coefficients (not reported), one of the stronger individual causal effects on the quality of governance.

The more important point is that Centripetalism affects a wide range of governance outcomes. Consequently, we must consider not simply the causal influence of this variable on a single outcome, but rather the cumulative effect of Centripetalism across all outcomes of substantive import, including, but not limited to, the eight outcomes explored in Table 2. We presume that constitutional institutions have moderate effects across an extraordinarily wide range of policies and policy outcomes. Unless one theorizes the full range of these causal effects, one misses the programmatic significance of constitutional factors on the quality of governance.

## CAUSAL MECHANISMS

Although governance is a well-established topic of investigation, there are few integrative approaches to this time-honored subject, and even fewer that subject their hypotheses to systematic, global empirical testing.<sup>12</sup> This paper has been motivated by the need to build general theory from heretofore discrete empirical findings. We should clarify, however, that in adopting a "macro" approach to political economy we do not intend to denigrate the more finely grained analyses that now populate this field. These studies are well crafted and generally quite informative. Even so, we see a need for an occasional synthesis. Without such synthesis, we face the danger that, as this field grows, we shall know more and more about less and less.

The theory of centripetalism is an attempt to put the loose pieces of this vast puzzle together in a unified framework and to articulate an alternative to the reigning paradigm. Results reported here suggest that prevailing models of democratic governance—most of which reiterate the verities of the decentralist model—may be mistaken. Institutions that fragment power and decentralize sovereignty are likely to compromise, not to bolster, the quality of governance in a democratic polity.

In this final section, we provide a very brief account of the causal mechanisms that might plausibly con-

nect constitutional institutions—unitarism, parliamentarism, and list-PR—with better governance outcomes. Given space limitations, we delineate, in skeletal form, only the most general and most salient features of this causal story, explored elsewhere in greater detail (Gerring and Thacker N.d.).

E. E. Schattschneider (1960) reminds us that every polity is biased in favor of some forms of popular participation, and against others. Political institutions can hardly play a neutral role in the organization of interests, ideas, and identities. Some activities will be "mobilized in," and others will be "mobilized out." The bias of a centripetal polity is toward highly institutionalized patterns of participation and decision making—specifically, strong political parties, corporatist-style interest representation, collegial decision making, and authoritative public administration.

First, centripetal institutions should encourage the formation of strong, centralized, and well-institutionalized political parties. Each of the three constitutional features of centripetalism helps to empower party leaders, disempower local leaders, and maintain the boundaries of each party's organization (Bowler, Farrell, and Katz 1999; Carey 2002; Carey and Shugart 1995).

Second, centripetal institutions should encourage a "corporatist" style of interest organization where interests are free from coercive state and party control but are (a) aligned with political parties, (b) coalesced into broad, "peak" associations, and (c) incorporated in a quasi-official capacity in the policymaking process. Preliminary research indicates that unitary, parliamentary, and list-PR institutions are likely to foster this distinctive style of interest representation, whereas decentralist arrangements should foster a more fragmented, free-floating, "pluralist," set of interests (Gerring and Thacker N.d.).

Third, centripetal institutions should help to promote a "collegial" (i.e., cooperative, consensual) style of decision making, as contrasted with the adversarial or individualistic styles of decision making common in centralist and decentralist polities. Collegial decision making is the norm wherever political power is vested in appointive or elective bodies that are engaged in regular face-to-face meetings (Baylis 1989, 7–8, 144; Sartori 1975). These include cabinets, cabinet committees, legislatures, legislative committees, party caucuses, commissions, regulatory bodies, and so forth: precisely the sort of institutions that list-PR electoral systems and parliamentary executives foster (Finer 1975; Longley and Davidson 1998). Parliamentarism offers equally important inducements to collegiality by virtue of its fusion of executive and legislative functions in the same body, the cabinet. In these circumstances, it is simply not possible for a serious and enduring division to spring up among the major actors: the prime minister, the cabinet, and the backbenchers. In a presidential system, by contrast, two separate institutions with overlapping powers yet different constituencies, (usually) different electoral cycles, and (often) a different partisan and ideological composition vie for power. For the most part, they are not on collegial terms

<sup>12</sup> Lijphart (1999) conducts a series of tests of his "consensus" theory of democratic governance. However, these tests are limited to 36 countries, a cross-sectional empirical format, and few control variables. A final difficulty is that Lijphart's tests include only one component ("executive-parties") of a two-dimensional theory; the other component ("federal-unitary") falls by the wayside. Huber, Ragin, and Stephens (1993) examine all OECD countries, an even smaller subset of democracies around the world, and focus only on one dependent variable, social welfare spending. This particular outcome, although useful, is a controversial measure of good governance because many writers associate good governance with small government (thus, we exclude aggregate spending from our battery of tests). Henisz (2000) includes a much broader range of cases and variables (though a much smaller purview of governance outcomes). However, the main index variable employed to measure the number of veto points ("Political Constraints") is opaque in its construction. Some of his choices seem to conflate institutional variables and policy outcomes: for example, the decision to measure judicial independence by the extent of "law and order" in a society.

with one another. Parliamentarism has an additional effect: the executive office itself—and its relations with the civil service—operates in a more collegial fashion in parliamentary systems than in presidential systems. In the latter, the executive is embodied in the person of the president. He, like the monarch, is the sole constitutional authority within the executive branch. In such an environment, it is easy to see why presidential executives tend to embody either a “hierarchical” model or one in which there is little formal organization at all (an “individualistic” model; Blondel and Manning 2002; Manning et al. 1999).

Finally, centripetal institutions should help to create an “authoritative” mode of public administration. This constitutional framework establishes a single principal (the cabinet) in charge of multiple agents holding distinct, nonoverlapping mandates. Centripetal polities feature clear lines of authority; thus, they tend to establish greater accountability between elected and appointive officials. Divided authority, by contrast, leads to mixed messages, overlapping jurisdictions, and rigid and detailed rules of procedure (“red tape”). Bureaucratic malfeasance is easily buried in the chaos or, if discovered, can be disavowed (“blame-avoidance”). Parallel institutions cannot hold other institutions accountable precisely because each institution is formally independent. Decentralized power structures also introduce coordination problems among political units wherever the actors are (a) multiple, (b) organizationally independent, (c) instilled with different perspectives and different organizational missions, and (d) empowered with an effective policy veto (Moe and Caldwell 1994).

We anticipate that each of these four institutions—strong political parties, corporatist-style interest representation, collegial decision making, and authoritative public administration—serves as a causal pathway running from centripetal constitutional institutions to good governance outcomes. Granted, these intermediate variables are difficult to measure, thus virtually impossible to test in a rigorous manner. Even so, a large body of case studies, small-N comparative studies, and theoretical work may be cited in support of the foregoing arguments (for further discussion, see Gerring and Thacker N.d.).

Taken together, the empirical and theoretical claims of this study have important ramifications for the ways that policymakers and constitution-makers conceive the task of constitutional engineering. In particular, they suggest that unitary polities may offer better governance than do federal arrangements, that parliamentary systems are superior to presidential systems, and that list-proportional electoral systems are better than winner-take-all or preferential-vote electoral systems, all else being equal. And they suggest that political leaders might look to devise other political institutions (not tested here) that successfully combine centralized authority with a broad-based inclusion of diverse interests, ideas, and identities. Good governance, we surmise, arises from institutions that pull toward the center, offering incentives to participate and disincentives to defect—voice, not vetoes.

## REFERENCES

- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2002. “Fractionalization.” Manuscript Harvard University.
- Banks, Arthur S. 1994. “Cross-National Time-Series Data Archive.” Center for Social Analysis, State University of New York at Binghamton. Binghamton, New York.
- Baylis, Thomas A. 1989. *Governing By Committee Collegial Leadership in Advanced Societies*. Albany: SUNY Press.
- Beck, Nathaniel. 2001. “Time-Series—Cross-Section Data: What Have We Learned in the Last Few Years?” *Annual Review of Political Science* 4: 271–93.
- Beck, Nathaniel L., and Jonathan N. Katz. 2001. “Throwing Out the Baby with the Bathwater: A Comment on Green, Yoon and Kim.” *International Organization* 55 (2): 487–95.
- Bentley, Arthur. 1908/1967. *The Process of Government*. Cambridge: Harvard University Press.
- Blondel, Jean, and Nick Manning. 2002. “Do Members of Government Do What They Say? Ministerial Unreliability, Collegial and Hierarchical Governments.” *Political Studies* 50 (3): 455–76.
- Bollen, Kenneth A. 1993. “Liberal Democracy. Validity and Method Factors in Cross-National Measures.” *American Journal of Political Science* 37: 1207–30.
- Bos, Eduard, My Vu, and Patience Stephens. 1992. “Sources of World Bank Estimates of Current Mortality Rates.” World Bank Working Paper Series 851. Washington: World Bank.
- Bowler, Shaun, David M. Farrell, and Richard S. Katz, eds. 1999. *Party Discipline and Parliamentary Government*. Columbus: Ohio State University Press.
- Bryce, James. 1905. *Constitutions*. New York: Oxford University Press.
- Buchanan, James M., and Gordon Tullock. 1962. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.
- Carey, John M. 2002. “Getting Their Way, or Getting in the Way?: Presidents and Party Unity in Legislative Voting.” Presented at the annual meetings of the American Political Science Association, Boston (August).
- Carey, John M., and Matthew Soberg Shugart. 1995. “Incentives to Cultivate a Personal Vote: A Rank Ordering of Electoral Formulas.” *Electoral Studies* 14 (4): 417–39.
- Carstairs, A. M. 1980. *A Short History of Electoral Systems in Western Europe*. London: Allen & Unwin.
- Cheibub, Jose Antonio. 1998. “Political Regimes and the Extractive Capacity of Governments: Taxation in Democracies and Dictatorships.” *World Politics* 50 (2) (April): 349–76.
- Cox, Gary W. 1990. “Centripetal and Centrifugal Incentives in Electoral Systems.” *American Journal of Political Science* 34: 903–35.
- Dahl, Robert A. 1956. *Preface to Democratic Theory*. Chicago: University of Chicago Press.
- Euromoney. 2004. “Country Risk Methodology.” Downloaded from Euromoney subscriber section, <http://www.euromoney.com/includes/magazine/print.asp>. (June 11).
- Finer, Samuel E., ed. 1975. *Adversary Politics and Electoral Reform*. London: Anthony Wigram.
- Gerring, John, Philip Bond, and William Barndt. N.d. “Democracy and Growth: A Historical Perspective.” Manuscript, Boston University. Forthcoming.
- Gerring, John, and Strom C. Thacker. 2004. “Political Institutions and Corruption: The Role of Unitarism and Parliamentarism.” *British Journal of Political Science* (April). 295–330.
- Gerring, John, and Strom C. Thacker. N.d. *Good Government: A Centripetal Theory of Democratic Governance*. Manuscript, Boston University. Forthcoming.
- Gerring, John, Strom C. Thacker and Carola Moreno. N.d., a. “Are Parliamentary Systems Better?” Manuscript, Boston University. Forthcoming.
- Gerring, John, Strom C. Thacker, and Carola Moreno. N.d., b. “Are Unitary Systems Better than Federal Systems?” Manuscript, Boston University. Forthcoming.
- Gerring, John, Strom C. Thacker and Carola Moreno. N.d., c. “Do Neoliberal Policies Kill or Save Lives? A Crossnational Analysis.” Manuscript, Boston University. Forthcoming.

- Gordon, Scott. 1999. *Controlling the State: Constitutionalism from Ancient Athens to Today*. Cambridge: Harvard University Press.
- Hart, Jenifer. 1992. *Proportional Representation. Critics of the British Electoral System, 1820–1945*. Oxford: Oxford University Press.
- Henisz, Witold J. 2000. "The Institutional Environment for Economic Growth." *Economics and Politics* 12 (1): 1–32.
- Herring, Pendleton. 1940. *The Politics of Democracy: American Parties in Action*. New York: Norton.
- Heston, Alan, Robert Summers, and Bettina Aten. 2002. *Penn World Table Version 6.1*. Center for International Comparisons at the University of Pennsylvania (CICUP). (October).
- Hirst, Paul Q. ed. 1989. *The Pluralist Theory of the State: Selected Writings of G.D.H. Cole, J.N. Figgis, and H.J. Laski*. London: Routledge.
- Howell, Llewellyn D. 1998. *The Handbook of Country and Political Risk Analysis*, 2nd ed. East Syracuse, NY: PRS Group.
- Huber, Evelyn, Charles Ragin, and John D. Stephens. 1993. "Social Democracy, Christian Democracy, Constitutional Structure and the Welfare State." *American Journal of Sociology* 99 (3) (November): 711–49.
- Humphreys, Macartan. 2005. "Natural Resources, Conflict, and Conflict Resolution: Uncovering the Mechanisms." *Journal of Conflict Resolution*, 49: 4, 508–37.
- Krueger, Anne O. 1995. *Trade Policies and Developing Nations*. Washington, DC: Brookings.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1999. "The Quality of Government." *Journal of Economics, Law and Organization* 15 (1): 222–79.
- Lieberman, Evan S. 2002. "Taxation Data as Indicators of State-Society Relations: Possibilities and Pitfalls in Cross-National Research." *Studies in Comparative International Development* 36 (4) (Winter): 89–115.
- Lijphart, Arend. 1999. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven: Yale University Press.
- Linz, Juan. 1994. "Presidential or Parliamentary Democracy. Does it Make a Difference?" In *The Failure of Presidential Democracy*, eds. Juan J. Linz and Arturo Valenzuela. Baltimore: The Johns Hopkins University Press.
- Longley, Lawrence D., and Roger H. Davidson, eds. 1998. *The New Roles of Parliamentary Committees*. London: Frank Cass.
- Lowi, Theodore. 1969. *The End of Liberalism*. New York: Norton.
- Manning, Nick, Naazneen Barma, Jean Blondel, Elsa Pilichowski, and Vincent Wright. 1999. *Strategic Decisionmaking in Cabinet Government: Institutional Underpinnings and Obstacles*. Washington DC: World Bank.
- Marshall, Monty. 1999. *Major Armed Conflicts and Conflict Regions, 1946–1997*. Dataset from CIDCM, University of Maryland. Obtained via the State Failure Task Force dataset, <http://gking.harvard.edu/data.shtml> (April 25, 2005).
- Marshall, Monty, G., and Keith Jagers. 2005. "Polity IV Project. Political Regime Characteristics and Transitions, 1800–2003." [www.cidcm.umd.edu/inscr/polity/](http://www.cidcm.umd.edu/inscr/polity/) (April 25, 2005).
- McConnell, Grant. 1966. *Private Power and American Democracy*. New York: Alfred A. Knopf.
- Mill, John Stuart. 1865/1958. *Considerations on Representative Government*, 3d ed., ed. Curren V. Shields. Indianapolis: Bobbs-Merrill.
- Moe, Terry M., and Michael Caldwell. 1994. "The Institutional Foundations of Democratic Government: A Comparison of Presidential and Parliamentary Systems." *Journal of Institutional and Theoretical Economics* 150: 171–95.
- Moon, Bruce E. 1991. *The Political Economy of Basic Human Needs*. Ithaca: Cornell University Press.
- Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–708.
- Nourret, Serge, ed. 1990. *Political Strategies and Electoral Reforms: Origins of Voting Systems in Europe in the 19th and 20th Centuries*. Baden-Baden: Nomos.
- North, Douglass C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History* 49: 803–32.
- Oates, Wallace E. 1972. *Fiscal Federalism*. Harcourt, Brace, Jovanovich.
- O'Donnell, Guillermo. 1999. "Horizontal Accountability in New Democracies." In *The Self-Restraining State: Power and Accountability in New Democracies*, eds. Andreas Schedler, Larry Diamond, and Marc F. Plattner. Boulder, CO: Lynne Rienner.
- Persson, Torsten, Gerard Roland, and Guido Tabellini. 1997. "Separation of Powers and Political Accountability." *Quarterly Journal of Economics* 112: 1163–1202.
- Piano, Aili, and Arch Puddington (eds.). 2004. *Freedom in the World 2004: The Annual Survey of Political Rights and Civil Liberties*. New York: Rowman & Littlefield.
- Ranney, Austin. 1962. *The Doctrine of Responsible Party Government: Its Origins and Present State*. Urbana: University of Illinois.
- Reiss, Julian. 2003. "Practice Ahead of Theory. Instrumental Variables, Natural Experiments and Inductivism in Econometrics." Unpublished manuscript, Centre for Philosophy of Natural and Social Science, London School of Economics.
- Riker, William H. 1982. *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. San Francisco: Freeman.
- Riley, James C. 2001. *Rising Life Expectancy: A Global History*. Cambridge: Cambridge University Press.
- Rodrik, Dani. 1995. "Comments." In *Trade Policies and Developing Nations*, ed. Anne O. Krueger. Washington, DC: Brookings, 101–111.
- Sartori, Giovanni. 1975. "Will Democracy Kill Democracy? Decision-Making by Majorities and by Committees." *Government and Opposition* 10 (Spring): 135–58.
- Sartori, Giovanni. 1976. *Parties and Party Systems: A Framework for Analysis*. Cambridge: Cambridge University Press.
- Schattschneider, E. E. 1960. *The Semi-Sovereign People*. New York: Holt, Rinehart, & Winston.
- Steinmo, Sven, Kathleen Thelen, and Frank Longstrech, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Cambridge: Cambridge University Press.
- Stepan, Alfred, and Cindy Skach. 1993. "Constitutional Frameworks and Democratic Consolidation: Parliamentarianism versus Presidentialism." *World Politics* 46: 1–22.
- Tiebout, Charles M. 1956. "A Pure Theory of Local Government Expenditure." *Journal of Political Economy* 64: 416–24.
- Truman, David B. 1951. *The Governmental Process*. New York: Alfred A. Knopf.
- Vanhanen, Tatu. 1990. *The Process of Democratization: A Comparative Study of 147 States, 1980–88*. New York: Crane Russak.
- Vile, M. J. C. 1967/1998. *Constitutionalism and the Separation of Powers*. Liberty Fund.
- Wooldridge, Jeffrey. 2002. *Introductory Econometrics: A Modern Approach*. Cincinnati, OH: South-Western College Publishers.
- World Bank. 2003. *World Development Indicators 2003*. CD-ROM.

# "Interpose Your Friendly Hand": Political Supports for the Exercise of Judicial Review by the United States Supreme Court

KEITH E. WHITTINGTON *Princeton University*

*The exercise of constitutional review by an independent and active judiciary is commonly regarded as against the interest of current government officials, who presumably prefer to exercise power without interference. In this article, I advance an "overcoming obstructions" account of why judicial review might be supported by existing power holders. When current elected officials are obstructed from fully implementing their own policy agenda, they may favor the active exercise of constitutional review by a sympathetic judiciary to overcome those obstructions and disrupt the status quo. This provides an explanation for why current officeholders might tolerate an activist judiciary. This dynamic is illustrated with case studies from American constitutional history addressing obstructions associated with federalism, entrenched interests, and fragmented and cross-pressured political coalitions.*

How do we account for judicial activism in a context in which judges are assumed to be favorably disposed toward a governing coalition's political agenda? It is relatively easy to understand why an institution like judicial review might be normatively appealing in the abstract and might be inserted into a constitutional scheme by politically detached draftsmen, for whom constitutional review might serve as an attractive enforcement mechanism to constitutional precommitments (Ackerman 1991; Elster 2000).<sup>1</sup> Similarly, current government officials who are fearful of losing power may attempt to build up judicial authority and entrench their allies in the courts in the hopes that judicial review will be used against future government officials (e.g., Ginsburg 2003; Moravcsik 2000; Ramseyer 1994). Government officials who expect to retain power, however, are less obvious supporters of constitutional review. Instead of building up judicial authority, they are likely to subvert it, and active judicial review may simply be a short-lived, transitional phenomenon that will be snuffed out once a political coalition consolidates its power over the government (Dahl 1957). Although a court with an accumulated stockpile of political capital with the general public might nonetheless be able to overcome hostile government officials in particular decisions (Caldeira 1986; Vanberg 2001), it seems likely that in time elected officials would be able to bring the judiciary into line. Does the judiciary sink into passivity at that point?

Though federal judges are protected by such securities as lifetime tenure and guaranteed salaries from political retaliation for their decisions, the judiciary as a whole is still vulnerable to politics (Ferejohn 1999).

Most routinely, the political appointments process creates regular opportunities for elected officials to bring the Court into line with political preferences (Dahl 1957; Stimson, Mackuen, and Erikson 1995). Despite the life-tenure of judges, a variety of legislative sticks are available to punish the Court for politically unpopular decisions. Court-curbing actions, by constitutional amendment, statute, or impeachment, have been frequently threatened over the course of American history, and often that threat has been sufficient to alter judicial behavior (Epstein and Knight 1998; Nagel 1965; Rosenberg 1992). Government officials can also limit the power of the Court by simply evading judicial edicts, which highlights the vulnerability of a judiciary that lacks, as Alexander Hamilton promised, both the executive sword and the legislative will (Hamilton 1961; Rosenberg 1991; Vanberg 2001).

Even in the American context, the maintenance of the judicial authority to interpret the Constitution and actively use the power of constitutional review is an ongoing political project. For "judicial activism," in the sense of the frequent constitutional invalidation of legislation and executive action, to be sustained over time, the courts must operate in a favorable political environment.<sup>2</sup> Judges must find reason to raise objections to government actions, and elected officials must find reason to refrain from sanctioning judges for raising such objections.

I consider the conditions under which judicial activism by a relatively friendly court may emerge and be sustained.<sup>3</sup> Given the global rise of the power of constitutional review and the persistent activism of the U.S. Supreme Court, it is important to understand the political supports for the exercise of judicial review.

Keith E. Whittington is Professor, Department of Politics, Corwin Hall, Princeton University, Princeton, NJ 08544 (kewhitt@princeton.edu).

I am grateful to the participants in the Center for American Political Studies seminar series at Harvard University, The Table, the Constitutional Theory conference at NYU Law School, and the Law and Politics workshop at Washington University, and to the anonymous reviewers for their helpful comments.

<sup>1</sup> Judicial review may also be a useful device for making "credible commitments" by current government officials to other powerful actors who would otherwise threaten their power (Landes and Posner 1975; Moustafa 2003; Weingast 1997).

<sup>2</sup> Although "judicial activism" is an ambiguous term of limited general utility, I employ it here in the specific sense of invalidation of legislative and executive action (see also Caldeira and McCrone 1982). As such, it connects with popular discourse about the courts and is consistent with a prominent dimension of common usage.

<sup>3</sup> James Rogers (2001) has likewise suggested an informational theory of judicial review by which legislators might rely on sympathetic courts to exercise the power of judicial review to correct inadvertent constitutional errors. It is unclear how politically important such a judicial function might be in practice (Whittington 2003), but it could work in complement with the friendly judicial review laid out here.

The existing normative and empirical literature on judicial independence and constitutional review largely emphasizes how judicial activism emerges when the judiciary is relatively unfriendly to the current legislative majority.<sup>4</sup> An emerging literature is concerned with showing how Supreme Court doctrine fits within goals and tensions within the broader political regime, however (e.g., Gillman 2002; Graber 1993; Pickerill and Clayton 2004; Tushnet Forthcoming). This emerging literature has observed that the exercise of judicial review often does not fit the "countermajoritarian" framework, but efforts to develop explanations for the emergence of judicial review are still in their initial stages. Here I suggest how structural characteristics of political systems such as the United States encourage cooperation between judges and political leaders to obtain common objectives. In particular, the Court assists powerful officials within the current government in overcoming various structural barriers to realizing their ideological objectives through direct political action. After sketching the logic of judicial review as a solution to the structural obstacles to direct political action, I consider three such obstacles in American politics—federalism, entrenched interests, and fragmented political coalitions—and illustrate how significant episodes of judicial review in the past have been consistent with this logic.

## JUDICIAL REVIEW BY AN ALLIED COURT

The establishment and maintenance of judicial review is a way of delegating some kinds of political decisions to a relatively politically insulated institution. This delegation aspect of judicial review drives the entrenchment thesis, as current political majorities attempt to insulate their policy preferences from future political majorities by empowering sympathetic judges who will endure through the electoral transition. This is only one of the potential uses to which such an institution may be put, however. Political majorities may effectively delegate a range of tasks to a judicial agent that the courts may be able to perform more effectively or reliably than the elected officials can acting directly.

It is well recognized that explicit or implicit "delegation" of political tasks to differently situated institutions and actors can be valuable in a range of political contexts (see generally, Voigt and Salzberger 2002). Legislative party leaders can solve collective action problems and protect the value of party labels (Cox and McCubbins 1993; Kiewiet and McCubbins 1991). Legislative committees can develop expertise and provide the information needed to make good policy (Krehbiel 1991). Central banks and independent judiciaries can allow legislators to credibly commit to policies valued by key constituencies (Landes and Posner 1975; Maxfield 1997). Interest groups can

develop cheap information on the performance of bureaucracies or the preferences of the electorate (Hansen 1991; McCubbins and Schwartz 1984). At the same time, it should be recognized that apparent legislative delegations may be better understood as the exploitation of available political resources and legislative weaknesses by other actors, such as executive branch officials, to enhance their own institutional position (Whittington and Carpenter 2003). Thus, we should be sensitive to the interaction between courts exploiting political opportunities and legislative leaders managing political risk.

The courts exercising a power of judicial review may be a vehicle for overcoming political barriers that hamper a governing coalition. There are two preconditions for this possibility to be reasonable. The first is that courts often be ideologically friendly to the governing coalition. Political majorities are unlikely to benefit from supporting courts that are ideologically divergent from them and are unlikely often to be able to work in tandem with them to achieve common political goals. There are reasons to believe that this precondition is often met in the American context, with the selection of individual judges (Dahl 1957), the departure of current judges (Spriggs and Wahlbeck 1995), the expansion of the judiciary as a whole (Barrow, Zuk, and Gryski 1996; De Figueiredo and Tiller 1996), and the structure of court jurisdiction (Gillman 2002) all facilitating the creation of a sympathetic judiciary. This is not to say that presidents and parties are never surprised by their judicial appointments or by judicial decisions, but merely that the Court often shares the constitutional and ideological sensibilities of political leaders.

The second precondition is that judicial review is actually useful to current political majorities. The usefulness to legislators of other judicial powers, such as the power to interpret statutes and enforce the law, is fairly evident. The utility of the power of judicial review to current legislators is less immediately evident, but it is easy to see once we note that judicial review may be used to void statutes passed by previous governing coalitions, thus displacing the current legislative baseline. When governing coalitions are unable or unwilling to displace the legislative baseline themselves, then the courts may usefully do this work for them. Those invested in the status quo have less to gain from judicial review (Graber 2000), and so judicial review is likely to be more useful to some political coalitions than others, depending in part on their substantive agenda and in part on the extent to which they have been able to define the status quo. Nonetheless, as is illustrated in the following, it is unrealistic to assume that only political actors currently out of power stand to benefit from an active judiciary.

We can expect that there will be additional supports for the active exercise of judicial review by an ideologically friendly judiciary to the extent that there are political barriers that hamper the realization of a governing coalition's agenda. In essence, allied elected officials would stand to benefit from an active judiciary if the ability of those elected officials to reach their preferred policy position on their own is limited.

<sup>4</sup> Even those who tend to assume that "successful constitutional judicial review" requires the acceptance of it by "other powerful political actors" nonetheless sometimes portray judicial review as itself undesired, "as an inevitable cost of getting [something else that] they want from courts" (Shapiro 1999, 210)



The resort to judges by displaced elected officials or minority interests is merely a special case of a larger class of cases in which political actors allied with the courts cannot control the legislative baseline. Political leaders who are still part of the governing coalition may nonetheless find their ability to implement their preferred policy hampered by difficulties other than simple electoral defeat. In a federal system, for example, ideological and partisan opponents may control policymaking jurisdictions that are insulated from direct national legislative control. In the context of heterogeneous and cross-pressured political coalitions, political leaders may be unable to mobilize legislative allies behind a given policy that nonetheless is viewed sympathetically by judicial allies.

Political leaders in such a situation will have reason to support or, at minimum, tolerate the active exercise of judicial review. In the American context, the presidency is a particularly useful site for locating such behavior. The Constitution gives the president a powerful role in selecting and speaking to federal judges. As national party leaders, presidents and presidential candidates are both conscious of the fragmented nature of American political parties and sensitive to policy goals that will not be shared by all of the president's putative partisan allies in Congress. We would expect political support for judicial review to make itself apparent in any of four fields of activity: (1) in the selection of "activist" judges, (2) in the encouragement of specific judicial action consistent with the political needs of coalition leaders, (3) in the congenial reception of judicial action after it has been taken, and (4) in the public expression of generalized support for judicial supremacy in the articulation of constitutional commitments.

Although it might sometimes be the case that judges and elected officials act in more-or-less explicit concert to shift the politically appropriate decisions into the judicial arena for resolution, it is also the case that judges might act independently of elected officials but nonetheless in ways that elected officials find congenial to their own interests and are willing and able to accommodate. Although Attorney General Richard Olney and perhaps President Grover Cleveland thought the 1894 federal income tax was politically unwise and socially unjust, they did not necessarily therefore think judicial intervention was appropriate in the case considered in more detail later (Eggert 1974, 101–14). If a majority of the justices and Cleveland-allies in and around the administration had more serious doubts about the constitutionality of the tax, however, the White House would hardly feel aggrieved. We should be equally interested in how judges might exploit the political space open to them to render controversial decisions and in how elected officials might anticipate the utility of future acts of judicial review to their own interests.

It should be emphasized that the possibility of friendly judicial review does not mean that the Court will simply do the bidding of political leaders. Politicians do not know with certainty what the justices will do if presented with a given piece of legislation. Although presidents may hope that the Court will act

in a given case, they may well be disappointed. When signing campaign finance reform, President George W. Bush virtually drew a roadmap of the statutory provisions that he hoped the Court would strike down, but a majority of the justices imposed only modest constraints on the congressional authority to regulate political campaigns (Bush 2002, 517; *McConnell v. Federal Election Commission* 2003). Striking down that statute might have won favor from the president who had signed it, but the Court merely behaved in the politically conventional manner by lending its legitimacy to the law.

At other times, the justices might well act on their own constitutional understandings even when those understandings are not shared by political leaders or when their expression is not desired. The political logic for such instances of unfriendly and unwelcome judicial review will have to be rather different from those described here. If the obstruction is relatively minor, as when the Court struck down Theodore Roosevelt's Employers' Liability Act as being drafted too broadly while indicating that the law's aims were constitutionally legitimate, then the Court's accumulated political capital might encourage leaders to simply yield to or work around the Court's rules (*Employers' Liability Cases* 1907; Pickerill 2004). If the obstruction is more serious, as when the Hughes Court blocked major components of the New Deal or when the early Warren Court extended the constitutional protections of suspected Communists, then the political reaction might be more severe and the strength of the Court's diffuse support might be tested. Not all episodes of judicial review take the collaborative form described here. The possibility of friendly judicial review, however, gives political leaders reason not only to tolerate the Court when it behaves in politically difficult ways but also to actively support the Court and help build a reservoir of public goodwill when it behaves in politically useful ways (Whittington *Forthcoming*).

I consider here three common barriers to successful action on ideological agenda items for political coalitions in American politics: federalism, entrenched interests, and coalitional heterogeneity. It should be noted that particular instances of judicial review may often involve more than one political logic. An instance of judicial review may well involve state action, for example, even when the structural obstacle of federalism is not the central political dynamic involved in the case. In each case, the central logic of the obstacle and how the exercise of judicial review may be useful for overcoming it is sketched out. In each instance, the Court is able to do what national political leaders are either constitutionally incapable of doing or politically unwilling to do themselves, and in doing so the Court runs with rather than against the interests of powerful political officials. An empirical illustration of this dynamic at work in significant episodes in American history is then provided. These cases are clearly not sufficient to indicate how much of the Supreme Court's exercise of judicial review can be explained in these terms, but they are sufficient to suggest that this dynamic has been a notable component of



the political support for judicial review in the United States and has been relevant to substantively important episodes of activism by the Court, thus expanding our conceptual toolkit for understanding the politics of judicial review.

## OVERCOMING FEDERALISM

Historically, the federal context has been an important one, perhaps the most important one, for generating support for the power of judicial review from other national government officials.<sup>5</sup> The Supreme Court has won the approval of national officials by imposing their shared constitutional agenda on recalcitrant state actors who hamper national political goals. Over the course of its history, the states have occupied more of the Court's constitutional attention than has the federal government, and the states have been the primary target of the power of judicial review. The Supreme Court has struck down state and local policies in well over 1,100 cases, but has rejected federal policies in just over 150 cases. Many of the most controversial political issues that have come before the Court have done so through cases involving the states. Despite the more recent celebration of the Court's review of Congress in *Marbury v. Madison*, the Court largely built its power of judicial review in the early decades of the U.S. government by acting against the states. Although the Court made few efforts to impose restrictions on the national government until after the Civil War, it struck down an average of six state statutes per decade in the early and mid-nineteenth century. In doing so, the Court found political advantage in upholding national supremacy, resolving interstate disputes, and securing the constitutional understandings favored by national political officials when those national officials could not act directly.

The fragmented American political system provides ample opportunities for national electoral minorities to nonetheless exercise political power. Particularly notable is the American federal structure, which allows ideological outliers and members of the out-party to consolidate and exercise governmental power over limited geographic jurisdictions. The independence of state and local governments from the national government is a source of ferment and resistance within the constitutional regime that national political officials might seek to establish. It was this very difficulty that led many advocates of constitutional reform in the 1780s to seek a stronger national government with a more effective capacity for disciplining subnational political actors (Banning 1995, 43–75; Rakove 1996, 51–53). Although delegates at the constitutional convention were unwilling to give Congress a direct, discretionary veto over state laws, they did draft the supremacy clause making explicit that the Constitution

trumped contrary state laws and implying the possibility of national judicial review of state actions.

James Madison was particularly moved, as many were, by the prospect of internecine violence and the promise of the judiciary as a way of securing union and preserving the peace (Deudney 1995; Hendrickson 2003). In the *Federalist*, Madison (1961, 245, 246) held up the Supreme Court "as the tribunal which is ultimately to decide" the limits of state and federal power. Every effort would be made to ensure the Court's impartiality and independence in resolving such issues, but regardless "some such tribunal is clearly essential to prevent an appeal to the sword and a dissolution of the compact." Decades later, Madison continued to affirm those early views despite the Court's doctrinal missteps. At least in those cases "not of that extreme character" the Court was "the authority constitutionally provided for deciding controversies concerning the boundaries of right and power" (Madison 1910, 9: 342–43). The alternative to such a "peaceful and effectual" system, he warned, was likely to be "the sword" (Madison 1910, 9: 348).

In this regard, John Marshall very much shared Madison's beliefs on the special role of the Supreme Court within the constitutional system. In his *McCulloch* decision in 1819, the Chief Justice observed that the controversy over Maryland's effort to use its taxing power to discourage the operation of the Bank of the United States within its borders pitted "a sovereign state" against the "legislature of the Union" and involved the "most interesting and vital parts" of the Constitution affecting the "great operations of the government." The issue, Marshall intoned, must be decided and must be "decided peacefully." If that peaceful settlement were to occur, "by this tribunal alone can the decision be made. On the Supreme Court of the United States has the Constitution of our country devolved this important duty" (*McCulloch v. Maryland* 1819, 400, 401). As the U.S. Attorney General requested in his arguments before the Court, the justices struck down the Maryland law. Marshall (1969, 212, 208) later elaborated in a pseudonymous defense of his opinion, for judges alone is "their paramount interest . . . public prosperity." Indeed, "if we were now making, instead of a controversy, a constitution, where else could this important duty of deciding questions which grow out of the constitution, and the laws of the union, be safely and wisely placed." The Court was not the first to interpret the Constitution's relevance to the Bank of the United States, but Marshall insisted that it should be the last. Although some Jeffersonians were unhappy with some of the language in Marshall's opinion, it echoed prominent voices among the National Republicans who dominated national politics after the War of 1812 and both former-president James Madison and the sitting administration of James Monroe quickly endorsed the decision and encouraged general compliance (Graber 1998, 256–57; Warren 1926, 1: 507–12). Though often remembered now as a deferential decision upholding congressional authority, in the context of the time *McCulloch* was decidedly activist, but the activism was directed against the states on behalf of the

<sup>5</sup> By explicitly laying aside judicial review of state legislation, Dahl (1957) made the Court seem far more passive than it has in fact been (Casper 1976). As discussed here, incorporating federalism into the political story of judicial review helps show how an active Court is still consistent with a politically responsive Court.

constitutional commitments of the national coalition (Whittington 2001a).

William Wirt, Monroe's respected attorney general, was an active force in building support for the Court during this period. Writing to President Monroe, Wirt dismissed the "few exasperated portions of our people" who, responding to "local irritations," favored "narrowing the sphere of action of that Court and subduing its energies." The "far greater number . . . wish to see it in the free and independent exercise of its constitutional powers, as the best means of preserving the Constitution itself." Indeed, Wirt judged that it "is now seen on every hand, that the functions to be performed by the Supreme Court of the United States are among the most difficult and perilous which are to be performed under the Constitution" (Kennedy 1850, 2:134). Arguing before the Court itself in 1824, the attorney general called on the Court to "interpose your friendly hand" and strike down New York's steamship monopoly. "It is the high province of this Court to interpose its benign and mediatorial influence" to "extirpate the seeds of anarchy" and stave off "civil war." So important was the Court in interposing the national will against the states that the constitutional framers would have deserved their "wreath of immortality" if they had done "nothing else than to establish this guardian tribunal, to harmonize the jarring elements of our system" (*Gibbons v. Ogden* 1824, 229). The Court acceded to the administration's request.

Even after the threat of intergovernmental violence receded, national officials have been no less concerned with curbing constitutional dissenters among the states. In concert with Republicans and conservative Democrats in Congress and the White House, the Court moved aggressively in the late nineteenth century, for example, to strike down state "legislative barriers ["to the consolidation of the national market"] almost as fast they were erected" (Bensel 2000, 324; see also, Kutler 1968). When the national "corporations uniformly fell back on their constitutional guaranties . . . [and] sought shelter behind the Constitution of the United States" from the ravages of various locally influential farmers' movements, the Court, after some initial hesitation, stood ready to extend constitutional protection to them (Adams 1875, 413). By the final decades of the nineteenth century, "the legislatures of the States . . . [had been made] subject to the superintendence of the judiciary" as the Court elaborated the economic liberties it found in the Constitution and the Fourteenth Amendment and talk of the "centralizing tendencies in the Supreme Court" was commonplace (Anonymous 1890, 521; Powers 1890, 389).

Although reformist elements made few inroads in the national government during the Gilded Age, they were able to set policy in a number of states. Conservatives called for the courts to intervene to stop the menace. In the preface of his celebrated treatise on the limits of the constitutional authority of the states, the young constitutional law professor Christopher Tiedeman (1886, viii) called for "a full appreciation of the power of constitutional limitations to protect private rights against the radical experimentation of

social reformers" who might regulate railroads, impair creditors, or burden out-of-state businesses. This view was echoed across the country by the increasingly organized and vocal legal profession and often found influence in the White House. Even as Populists were ramping up their criticisms of the Court and the power of judicial review, Republican President Benjamin Harrison suggested a centennial celebration of the Supreme Court. The 1890 event, presided over by former President Grover Cleveland and sponsored by the New York Bar Association, featured Justice Stephen Field (1890, 367), who had been selected by his colleagues to deliver the message, emphasizing the "imperative duty of the court to enforce with a firm hand every guarantee of the constitution." A few years later, the American Bar Association organized a nationwide centennial celebration of the appointment of Chief Justice John Marshall to the Court, which became an occasion to celebrate the power of the courts to interpret and enforce the Constitution, the American innovation that threw off "the doctrines and theories engendered by the French Revolution—the supreme and uncontrollable right of the people to govern" (Dillon 1903, 1: xviii).

The Court has often used the power of judicial review to bring the states into line with the nationally dominant constitutional vision. In his comprehensive analysis of state statutes and constitutional provisions invalidated by the Supreme Court from the Jacksonian era through 1964, John Gates (1987, 260) found that the Court was particularly likely to act against "states whose partisan character is different from the dominant majority on the Court or from regions which evidence ideological incongruence between the state and national party organizations." In the late nineteenth and early twentieth centuries, judicial review by a conservative Court was primarily exercised against "regions where populism [and later progressivism] had made strong inroads" (Gates 1992, 67). In the mid-twentieth century, invalidated state laws emerged mostly from Republican states and the ideologically isolated South. As Michael Klarman (1996) and Lucas Powe (2000) have detailed, the Warren Court primarily targeted those states and interests who were resistant to national cultural and political trends. Political losers at the national level can often pursue their constitutional and policy proclivities in various state governments, but throughout its history the Supreme Court, with the encouragement of national leaders, has stood ready to "expand the scope of conflict" by pulling those policies back into the national arena for ultimate resolution (Schattschneider 1975).

## OVERCOMING ENTRENCHED INTERESTS

The American political system is fragmented horizontally within governments as well as vertically between layers of government. This fragmentation—across branches, across legislative chambers, and within legislative chambers—frequently obstructs those seeking to alter the status quo. Majority parties in the United States can rarely exercise the kind of policymaking

power exerted by governing coalitions in unitary, majoritarian political systems. Entrenched interests can often frustrate reform and can benefit from a powerful status-quo bias of American lawmaking. Coalition leaders who might prefer to embark on an ambitious programmatic agenda may only achieve partial success in the legislature. Just as presidents sometimes turn to their unilateral powers to make policy initiatives in order to circumvent legislative obstructions, so the courts can be a useful alternative vehicle for reform even for those who are part of the majority coalition. Clearly this happens in the statutory realm (Frymer 2003; Melnick 1994), but it can happen in the constitutional realm as well. In what Michael Klarman (1997) characterizes as "majoritarian judicial review," the judiciary can assist members of the political majority in dislodging entrenched political actors and interests. The same gridlock that hampers positive action by elected officials, however, also constrains their responsiveness to judicial decisions, facilitating judicial action that can count on the backing of well-placed elected officials.

The famed legislative apportionment decision of 1962 is an example of the Court cutting through the "political thicket" (*Colegrove v. Green* 1946, 556). Chief Justice Earl Warren (1977, 306) later regarded *Baker v. Carr* as "the most important case of my tenure on the Court." As governor of California, Warren (310) had contributed to the preservation of malapportioned and gerrymandered legislative districts, which he later admitted "was frankly a matter of political expediency." "But I saw the situation in a different light on the Court. There, you have a different responsibility." From that perspective, he came to believe that he "was just wrong as Governor" (Schwartz, 411). The Court's willingness to intervene in the field was an abrupt departure from the traditional understanding of apportionment being a legislative and deeply political prerogative, but it was a departure that was being urged on the Court by programmatic liberals in and around the White House. Often portrayed as an instance of the Court simply acting on behalf of popular majorities, legislative reapportionment was the specific project of liberal Democrats who had long chafed at the legislative obstacle posed by entrenched conservatives.

Others on the Court shared Warren's sense of the momentous significance of the case, but for quite different reasons. A bitter dissenter in the case, Frankfurter thought the decision was "bound to stimulate litigation by doctrinaire 'liberals' and the politically ambitious" that could only damage the Court in the long run (Schwartz 1983, 413). His ally John Marshall Harlan agreed and appealed to the swing justices not to open the door to such cases in which partisan politics and interest were so much on the surface. "Today," he noted, "state reapportionment is being espoused by a Democratic administration; the next time it may be supported (or opposed) by a Republican administration. Can it be that it will be only the cynics who may say that the outcome of a particular case was influenced by the political backgrounds or ideologies of the then members of the Court . . . ?" (Schwartz, 414). But Congress, Warren countered, had already pushed the justices into serv-

ing as "the referee" in state elections (Schwartz, 416). Justice Tom Clark had initially planned to write dissent in the case, emphasizing that nonjudicial remedies were available to address the malapportioned districts in Tennessee that were immediately at issue. After conducting the research for his opinion, however, Clark had to report to Frankfurter that he had changed his mind and would be joining the majority, "I am sorry to say that I cannot find any practical course that the people could take in bringing this about except through the Federal courts" (Schwartz, 423). Solicitor General Archibald Cox had emphasized the same point in his oral arguments as a friend of the Court, "Either there is a remedy in the Federal court or there is no remedy at all" (Special to *The New York Times* 1961, 25), and it figured prominently in the formal opinions of the justices (*Baker v. Carr* 1962, 248, 258-59).

The Court's willingness to extend constitutional principles to cover legislative apportionment was welcomed by liberals, who had long favored reapportionment as a means for reaching other programmatic goals but they had been stymied in the political process. The New Deal had pulled urban voters firmly into the Democratic coalition, and the malapportionment of the era overwhelmingly favored more conservative rural voters over more liberal urban voters. After Roosevelt's initial landslide victory, the *Nation* crowed, "For seventy-five years the Republicans have dominated the Northern and Eastern States through rotten-borough provisions in the State constitutions. . . . [but now] the day of retribution has come" (Welsh 1932, 523). But the day had not yet come, and a decade later it could only complain, "[T]he present gerrymandering of state districts amounts to supporters of the New Deal being denied equal voice with its opponents" (Neuberger 1941, 127).

Both the constitutional principle and the political consequences of judicial intervention were in line with the liberal regime. In the last years of the Eisenhower administration, Anthony Lewis (1958, 1059) of *The New York Times* had prominently pointed to the federal courts as the only institution politically capable of correcting "this growing evil of inequitably apportioned legislative districts," given the "virtually insurmountable, built-in obstacles to legislative action," and he exhorted the judges to take the lead. "A vacuum exists in our political system; the federal courts have the power and the duty to fill this vacuum." Taking a cue from the Supreme Court's boldness in *Brown*, federal district judge Frank McLaughlin, a Truman appointee and former New Deal congressman, declared that legislative inaction on reapportionment in Hawaii had gone on for too long; "The time has come, and the Supreme Court has marked the way, when serious consideration should be given to a reversal of the traditional reluctance of judicial intervention in legislative reapportionment. The whole thrust of today's legal climate is to end unconstitutional discrimination. It is ludicrous to preclude judicial relief when a main-spring of representative government is impaired" (*Dyer v. Abe* 1956, 226). While still a senator preparing for his presidential run, John F. Kennedy (1958, 37,

38) had published a magazine article calling legislative apportionment "deliberately rigged" and "shamefully ignored"; the only result of "this basic political discrimination," he argued, was the "frustration of progress." By then, the *Nation* could see the possibility of a "civil-liberties battle" over legislative apportionment being fought in the courts, and liberal interest groups such as the AFL-CIO, American Civil Liberties Union, and Americans for Democratic Action were early participants in apportionment litigation (Cortner 1970; Fleming 1959, 26). Even as friends of the Kennedy administration such as James MacGregor Burns (1963, 1) bemoaned the "old cycle of deadlock and drift" that killed "most of Mr. Kennedy's bold proposals," the *Nation* pointed to malapportionment as the linchpin of the conservative coalition's legislative power and encouraged the courts to pull it out (Lindsay 1962, 208). Doing so was expected not only to aid Democrats over Republicans but also pointedly to strengthen the hand of liberal Democrats at the expense of conservative Democrats.

The Kennedy electoral campaign concentrated on the urban vote, and once in the White House, the administration for the first time encouraged the Court to intervene in legislative apportionment in the case of *Baker v. Carr* and voiced its support after that favorable decision was announced. The Kennedy's had forced the reluctant Archibald Cox to argue the case before the Court. Upon release of the Court's decision, Attorney General Robert Kennedy immediately hailed it as a "landmark in the development of representative government" and observed that "the democratic process has been distorted," requiring an "effective judicial remedy" (Special to *The New York Times* 1962, 1). Publicly, the president endorsed the Court's decision and reminded the American people that the administration had in fact encouraged it. "Quite obviously," John Kennedy (1963, 274) asserted, "the right to fair representation and to have each vote counted equally is, it seems to me, basic to the successful operation of a democracy." It had been "impossible for the people involved to secure adequate relief through the normal political processes." Although it was the "responsibility of the political groups to respond to the need," when no relief was forthcoming "then of course it seemed to the Administration that the judicial branch must meet a responsibility." Privately, he elaborated to former Secretary of State Dean Acheson, "the legislatures would never reform themselves and that he did not see how we were going to make any progress unless the Court intervened" (Schwartz 1983, 425). Administration officials subsequently claimed credit for winning the result in *Baker*, and the Kennedy Justice Department remained active in subsequent reapportionment litigation (Sowell 1992, 383-84). Two years later, in another reapportionment case, Harlan complained, "these decisions give support to a current mistaken view of the Constitution and the constitutional function of this Court. This view, in a nutshell, is that every major social ill in this country can find its cure in some constitutional 'principle,' and that this Court should 'take the lead' in promot-

ing reform when other branches of government fail to act" (Reynolds v. Sims 1964, 624). For sympathetic political leaders, this view might have been current, but it was hardly politically mistaken. From the White House down, liberals turned to the Court in order to displace entrenched conservative legislators who could not be defeated by other means, and they contributed to the political and intellectual climate that would lend support and legitimacy to the Court taking that unprecedented step.

## OVERCOMING FRACTIOUS COALITIONS

American political parties are often fractious coalitions, and party unity may come at the price of substantial policy compromise. For the leaders of factions within the governing party, judicial review may offer the means for continuing the intracoalitional disagreement and potentially for undoing the compromises that had to be made in the political and legislative arenas. The backstop of friendly judicial review may smooth the legislative relations of the members of fractious political coalitions while providing some measure of additional security for the central commitments of party leaders and presidents. Judicial invalidation of even recent federal law will not necessarily be unwelcome by political leaders.

One of the more controversial exercises of judicial review in the nineteenth century—the invalidation of the federal income tax in 1895—fits this description. When Republicans controlled the federal government during the Civil War, they adopted many of the economic policies of their Whig predecessors, including the protective tariff. The protective tariff soon became a key plank in the Republican platform, and the Republicans kept duties on imported goods high whenever they held power until their conversion to free trade after the Second World War. The Democrats had been equally committed to free trade since the Jackson presidency, and when Grover Cleveland regained the White House for the Democrats, he railed against the protective tariff as injurious to consumers and an example of government corruption. When the federal government finally fell under unified Democratic control after the 1892 elections, Cleveland made tariff reform the centerpiece of his second term of office.

In the midst of economic depression and growing budget deficits, lowering tariffs was a tough sell. Nonetheless, Cleveland staked the future of the party on it and was personally active in designing the reform and pushing it through Congress. House Ways and Means Committee Chairman William Wilson, working closely with the president, immediately began negotiating tariff reform at the opening of the Fifty-third Congress. Despite presidential support and party ideology, however, many newly elected Democratic congressmen from manufacturing districts were loath to reduce import duties, while still others worried that significant tariff reform would not be consistent with a balanced budget without the addition of some other tax. To calm these latter concerns, Cleveland had endorsed the inclusion of a temporary "small tax upon incomes

derived from certain corporate investments" that could be lifted as soon as the fiscal climate improved, but the administration had earlier rejected efforts to include a personal income tax in the tariff bill (Richardson 1908, 9:460; Summers 1953, 152–86). This was not enough to win a majority, and the Republicans and Populists combined to deny the Democrats a functioning quorum. The Populists and Populist-leaning Democrats in the House were pivotal to the passage of any significant tariff reform, but the price of their cooperation was the inclusion of their income tax measure in the tariff bill. Over Wilson's objections, the Democratic caucus took the deal as the only way to salvage the president's program. Despite delaying motions of New York Democrats, who declared that "we stand here with the patron saints of Democracy, the apostles who have laid down the law of the party for 100 years... [and] declared internal taxation abominable," the majority of the Democrats in the House joined with the Populists to bundle the two measures and pass the whole (*The New York Times* 1894, 6). The situation was even worse in the Senate, where even more compromises had to be made on duty rates to keep a majority together.

President Cleveland was hardly satisfied with the results of the legislative negotiations. Despite his own misgivings, he was convinced that the bill "is so interwoven with Democratic pledges and Democratic success that our abandonment of the cause of the principles upon which it rests means party perfidy and party dishonor" (Cleveland 1933, 355). Although the amended bill fell well short of what they had wanted, Cleveland (357) rationalized to Wilson, "You know how much I deprecated the incorporation in the proposed bill of the income tax feature. In matters of this kind, however, which do not violate a fixed and recognized Democratic doctrine, we are willing to defer to the judgment of a majority of our Democratic brethren. I think there is general agreement that this is party duty," a duty that was all the more pressing when it was recognized that "a quick and certain return of prosperity waits upon a wise adjustment" to the tariff. Even though the president had strained to ensure the passage of the bill into law, he could not bring himself to sign such inadequate legislation. The Tariff Act of 1894 became law without the president's signature just a few months before the midterm election, but it was not enough to prevent the Democrats from being routed in both chambers of Congress. Months before the Republican majorities assembled in the Fifty-fourth Congress, however, the Supreme Court struck down the income tax provisions of the Tariff Act. When the Republicans regained the White House two years later, tariff rates were again adjusted upwards.

The income tax was harshly denounced as a purely sectional and class measure, and indeed it was. Nebraska Representative William Jennings Bryan, the emerging leader of the populist wing of the Democratic Party, was a primary sponsor of the amendments, and its support came almost exclusively from legislators from the South and West. The 2% tax on all personal income over \$4,000 was a significant symbolic shift from the traditional sources of federal revenue

and was expected to fall primarily on the residents of only four states (New York, New Jersey, Pennsylvania, and Massachusetts). Two of these states (New York and New Jersey) happened to also be important swing states in Gilded Age presidential elections, and New York in particular was essential to Democratic Electoral College calculations. It was the centrality of New York that led to reformist New York Governor Grover Cleveland's own Democratic presidential nomination in 1884, 1888, and 1892 and the integration of the Mugwumps (a breakaway group of Republican professionals and businessmen centered in New York) into the Cleveland coalition (James 2000, 42–56). Democratic New York Senator David Hill warned his populist colleagues, "The times are changing; the courts are changing, and I believe that this tax will be declared unconstitutional. At least I hope so" (*Congressional Record* 1894, 6637). The business community in New York was apoplectic over the income tax. Although some in the New York City press labeled it a Cleveland tax, his allies defended the president as an opponent of the tax and a victim of the populists (*The New York Times* 1896, 4).

Immediately upon its passage, a group of New York businessmen sponsored a collusive suit between a company and a stockholder to put the constitutionality of the income tax before the Court. The administration dutifully defended the constitutionality of the tax, calling on the Court to respect Federalist-era precedent and the appropriate sphere of legislative discretion over the proper exercise of the taxing power (*Pollack v. Farmers' Loan and Trust* 1895a, 502, 513). But the Court first struck down the tax on income from real estate and state and local bonds, and a month later a narrow majority struck down the rest. Cleveland-appointed Chief Justice Melville Fuller wrote both opinions striking down the provisions as violating basic constitutional efforts "to prevent an attack upon accumulated property by mere force of numbers" (*Pollack v. Farmers' Loan and Trust* 1895a, 583). Among the dissenters, Republican John Marshall Harlan was offended not least by the Court's willingness to undo the legislative compromise while leaving the tariff reduction still standing; "every one knows, the act never would have passed" without the income tax provisions (*Pollack v. Farmers' Loan and Trust* 1895b, 684).

The decision set off great rejoicing in some quarters, as *The New York Times* (1895b, 4) crowed that, although "enacted by a Democratic Congress," the tax was "not Democratic in theory or policy, and... the method of constitutional interpretation that has guided the Supreme Court in destroying them is one of the fundamental doctrines of the Democratic Party. The rendering of this opinion is an event of the utmost importance to that party." The decision also set off enormous criticism of the Court, led by Bryan who routed the Cleveland forces to capture the Democratic nomination the next year, but the president refrained from adding to the din and his loyalists in a breakaway party convention denounced Bryan for his attacks on judiciary (Stephenson 1999, 107–28). When income-tax dissenter Howell Jackson died just months after

the decision, Cleveland replaced him with conservative New York corporate attorney Rufus Peckham, whose nomination the president first cleared through Senator Hill. Of course, if Bryan had won the elections of 1896 the conservative-leaning Court might well have faced some difficulties. As it was, however, both conservative Democrats and the Republicans welcomed the Court's intervention and supported its increasing willingness to exercise the power of judicial review. As the Court prepared for reargument on the income tax, the Cleveland-allied *New York Times* (1895a, 4) expressed the sentiment of the ultimate victors when it editorialized that striking down the tax should be understood less as "magnifying the function of the Supreme Court" than as "resuming a function that had been to some extent abandoned, and with unfortunate, with really deplorable, results."

A century later, President Bill Clinton was similarly forced to swallow a disagreeable amendment in order to get a significant legislative package through Congress, and the subsequent exercise of judicial review can likewise be understood to have been friendly to the sitting administration. In February 1996, the president finally signed the Telecommunications Act, the most important telecommunications reform since the New Deal and an administration priority from the beginning of Clinton's term of office. Clinton (1997, 186) marked the occasion by traveling to the Library of Congress on Capitol Hill to sign a law that he promised would unleash the "free flow of information." He praised its potential "to build our economy . . . , to bring educational technology into every classroom, and to help families exercise control over how media influences their children" (Clinton, 127). The last was in recognition of the legislation's requirement of the "V-chip," the administration's favored technological fix to sex and violence on television. The president did not mention another high-profile element of the law, the Communications Decency Act, which the Justice Department would soon be defending in court.

The Communications Decency Act (CDA) was a last-minute amendment on the floor of the Senate to the telecommunications reform bill. Democratic Senator James Exon of Nebraska had originally introduced the measure in February 1995 to extend "the standards of decency which have protected telephone users to new telecommunication devices" (*Congressional Record* 1995, 3203). As the Senate neared final deliberations on the telecommunications bill, Exon and Republican Senator Daniel Coats offered a revised version of the CDA as an amendment. With lurid photos downloaded off the Internet available on his desk for his colleagues to view, Exon quickly won a lopsided vote to include the CDA in the reform bill. The Department of Justice and the Clinton administration had repeatedly voiced their opposition to the measure, judging it both unworkable and unconstitutional, but as Senator Orrin Hatch complained of the Senate vote, "It's kind of a game, to see who can be the most against pornography and obscenity . . . It's a political exercise" and the administration was unable to prevent its addition to the bill (Andrews 1995,

D6). The House of Representatives had already passed the reform bill with the administration's preferred indecency provision calling for the Justice Department to study the issue, and Speaker Newt Gingrich had denounced the Exon proposal as unconstitutional. After Senate passage, however, the Clinton administration relented, concluding, according to a senior administration official, "No way are you going to get yourself in a position where the president isn't willing to go as far as a Democratic senator in restricting child pornography on the Internet" in an election year (Weisberg 1996). It was initially hoped that the Senate's amendment would be excised in the privacy of the conference committee, but in a surprise victory for social conservatives the conference narrowly voted to adopt the Senate's language (Bryant and Plotnikoff 1996). At the same time, however, the conference did entrust enforcement to the Department of Justice (rather than the Federal Communications Commission) and provide for expedited judicial review of its indecency provisions. The president announced that he would not allow the inclusion of the CDA to hold up telecommunications reform, and with political attention now focused on it the Justice Department pledged to defend the measure "so long as we can assert a reasonable defense consistent with Supreme Court rulings in this area" (Schwartz 1996, A8).

The courts agreed with what the Justice Department told Congress rather than with what it said in its legal briefs. After a special three-judge panel struck down the CDA as unconstitutional in the summer of 1996, Clinton (1997, 906) affirmed that "I remain convinced, as I was when I signed the bill, that our Constitution allows us to help parents by enforcing this act," but said that the Justice Department would be responsible for a decision as to whether to appeal and trumpeted the administration's support for filtering software to block "objectionable materials." The administration quickly concluded that it would be politically costly not to appeal, however, and the Supreme Court struck down the provision in *Reno v. American Civil Liberties Union* (1997), severing it from the Telecommunications Act. The White House issued a statement reemphasizing its commitment to protecting children from inappropriate material and announcing plans for a conference to study blocking technology similar to the V-chip (Clinton 1998, 829). Exon lamented the Court's decision from his retirement in Nebraska, while his local paper hailed his "good try" (Knapp 1997; *Omaha World Herald* 1997).

## OVERCOMING CROSS-PRESSURED POLITICAL COALITIONS

There are some issues that politicians cannot easily handle. For individual legislators, their constituents may be sharply divided on a given issue or overwhelmingly hostile to a policy that the legislator would nonetheless like to see adopted. Party leaders, including presidents and legislative leaders, must similarly



sometimes manage deeply divided or cross-pressured coalitions. When faced with such issues, elected officials may actively seek to turn over controversial political questions to the courts so as to circumvent a paralyzed legislature and avoid the political fallout that would come with taking direct action themselves. As Mark Graber (1993) has detailed in cases such as slavery and abortion, elected officials may prefer judicial resolution of disruptive political issues to direct legislative action, especially when the courts are believed to be sympathetic to the politician's own substantive preferences but even when the attitude of the courts is uncertain or unfavorable (see also, Lovell 2003). Even when politicians do not invite judicial intervention, strategically minded courts will take into account not only the policy preferences of well-positioned policymakers but also the willingness of those potential policymakers to act if doing so means that they must assume responsibility for policy outcomes. For cross-pressured politicians and coalition leaders, shifting blame for controversial decisions to the Court and obscuring their own relationship to those decisions may preserve electoral support and coalition unity without threatening active judicial review (Arnold 1990; Fiorina 1986; Weaver 1986). The conditions for the exercise of judicial review may be relatively favorable when judicial invalidations of legislative policy can be managed to the electoral benefit of most legislators. In the cases considered previously, fractious coalitions produced legislation that presidents and party leaders deplored but were unwilling to block. Divisions within the governing coalition can also prevent legislative action that political leaders want taken, as illustrated in the following case.

This complicated dynamic can be illustrated through the consideration of Democratic strategies for dealing with the Court and racial civil rights in the 1950s. For Democrats, civil rights fell along the central fault line of their existing legislative and electoral coalition, dividing White Southern Democrats from more liberal Northern Democrats. Both Black voters in the North and White voters in the South were increasingly regarded as potentially pivotal in determining the control of the White House, but they put conflicting demands on presidential candidates. The Court as a policymaker was a potential strategic resource for overcoming a fragmented coalition and achieving policy outcomes greatly desired by some constituents. At the same time, the independence of the judiciary from explicit political control allowed politicians to distance themselves from judicial actions greatly disliked by other constituents, allowing politicians to roll with the judicial punches rather than having to retaliate against them.

For liberals during the Roosevelt and Truman administrations, racial civil rights suffered from a gridlock problem arising from within the Democrats' own electoral coalition. Decades of political neglect and the Great Depression tore the Black vote loose from the party of Lincoln. As Blacks continued to migrate north and became an important part of the voting constituency of Northern Democrats, Black civil rights became an increasingly salient issue for Northern liberals and national party leaders. Nonetheless, the pivotal

role of Southern Democrats in the New Deal legislative and electoral coalition stymied progress on the issue. By 1940, the Roosevelt administration had recognized the importance of the Black vote in the North, but rebuffed the NAACP so as not to risk higher priority agenda items (McMahon 2003; White 1948, 169–70). Hubert Humphrey rose to national prominence in the 1940s stumping for a "real, liberal Democratic Party" that would take action on Black civil rights and excommunicate Southern conservatives (Delton 2002, 120). Meanwhile, Truman was famously advised that the "Northern negro voter today holds the balance of power in Presidential elections" and that it was "inconceivable" that the South would revolt no matter how far to the left the administration leaned (Rowe 1995, 36, 30). In the election year of 1948, Truman (1964, 122) fruitlessly explained to Congress that the duty to secure civil rights "is shared by all three branches of the Government" and took some unilateral actions of his own. This proved to be enough to provoke Strom Thurmond's "Dixiecrat" revolt, which eventually stole 39 electoral votes from Truman in the general election. Though Truman won a surprising victory in 1948, the Dixiecrat scare hung over the Democratic Party for more than a decade.

In its second term, the Truman administration itself took a different tack on civil rights. Though "black activists and their white liberal allies from the programmatic wing of the Democratic party . . . were determined to press their cause even at the risk of disrupting the unity of the national party," others were centrally concerned with coalition maintenance (Sundquist 1983, 354). "Programmatic" advances would have to be accomplished through safer means. In public Truman largely dropped the issue, but his aides shifted resources into the Justice Department and sketched out a litigation strategy that would "offset the legislative defeats" (Berman 1970, 166). In its last years in office the administration filed briefs urging the Court to overthrow Jim Crow, and when stumping in Harlem for the 1952 Democratic ticket Truman (1968, 798) highlighted the actions that the administration had urged the Supreme Court to take.

Truman's Democratic successors were determined to downplay the civil rights issue. In 1952, Adlai Stevenson emerged as "the man most likely to hold together the liberal-labor-Southern coalition that Franklin D. Roosevelt built," though Black Democratic convention delegates walked off the floor when Alabama Senator John Sparkman was selected as the vice-presidential candidate (Reston 1952, 1). After *Brown* raised the stakes on civil rights, Stevenson remained insistent in 1956 that "where principle and unity conflicted in this matter, he was bound to stand by unity." Though he pledged that he would "act in the knowledge that law and order is the Executive's responsibility" and that it was "the sworn responsibility of the President of the nation to carry out the law of the land" as declared by the Court, he worked to keep the party from explicitly endorsing the *Brown* decision (Martin 1977, 302, 317). Stevenson's advisors initially assured him that the Court in *Brown* had ended civil rights as a political issue, but later changed their minds



and raised the specter of another Dixiecrat revolt but of "considerably greater magnitude" (Gillon 1987, 97; Martin 1979, 125). Pulled by both sides, Stevenson wailed in frustration during the 1956 primaries, "I had hoped the action of the Court and the notable record of compliance . . . would remove this issue from the political arena" and complained that the Eisenhower administration was not doing enough to make the issue go away faster (Martin 1977, 266).

In 1960, the Kennedy brothers likewise feared that becoming entangled in the civil rights issues would cost the party more votes than it would gain (Frymer 1999). Though approving the inclusion of a civil rights plank in the party platform, the Kennedy administration was determined not to "endorse a frontal assault against the segregation system" and when action was necessary "kept the president in the background, and stressed the need to uphold the law, rather than the moral right of blacks to use desegregated facilities" (Matusow 1984, 64, 74). The Justice Department advised citizens that civil rights were "individual," "private," and "personal" and to be pursued in court with their own attorneys (Marshall 1964, 50). Only when national and international public opinion turned decisively against Southern violence in 1963 did the president embrace civil rights as a "moral issue . . . as clear as the American Constitution" (Kennedy 1964, 469).

Although national party leaders ducked the issue, other Democratic politicians were free to play to their own local constituencies. In the aftermath of the *Brown* decision, Hubert Humphrey of Minnesota rushed to praise the Court for taking "another step in the forward march of democracy," while Dennis Chavez of New Mexico proclaimed that it "meets with my entire thinking and approval" (Albright 1954, 2). Northern congressional Democrats feared that in the wake of *Brown* "Republicans will move in on their once vast minority following" and found stronger appeals on the civil rights issue electorally essential (Albright 1956, E1). While party activists such as Joseph Rauh of Americans for Democratic Action proclaimed that the "Supreme Court has pointed the way for the future," the 1956 convention under Stevenson's watchful eye only recognized in the very last plank of its platform that "the Supreme Court of the United States as one of the three Constitutional and coordinate branches of the Federal Government [was] superior to and separate from any political party" and its decisions were "part of the law of the land" (Martin 1979, 150).

The reaction of Southern politicians was, of course, intense, including most famously the "Southern Manifesto" signed by most federal legislators from the Southern states (but the Speaker of the House and the Senate Majority Leader, both of Texas, were not asked to sign). Even so, the Manifesto limited itself to encouraging only "all lawful means to bring about a reversal of the decision," a restraint that both Stevenson and President Dwight Eisenhower praised. What the *Washington Post* called Southern "moderates," also notably national Democratic leaders heavily cross-pressured by their local constituencies, carefully shifted the blame for the federal government's new stance on civil rights while refraining from subverting judicial

review as such. Russell Long emphasized, "Although I completely disagree with the decision, my oath of office requires me to accept it as law. Every citizen is likewise bound by his oath of allegiance to his country" (Albright 1954, 2). Liberal Tennessee senator and presidential hopeful Estes Kefauver, under pressure from segregationists, explained to home state voters that his hands were tied by the Court, "There is not one thing that a member of the United States Senate can do about that decision—and anyone who tells you that he's going to do something about it is just trying to mislead you for votes" (Special to *The New York Times* 1954, 60). Richard Russell, also a Democratic presidential aspirant, went further and tried, in the *Post's* estimation, "to pin responsibility for the decision directly on the Republican administration," complaining that "the Supreme Court is becoming the political arm of the executive branch." Eisenhower's attorney general, Russell surmised, was intervening with "pressure groups" while the Court "supinely transposes the words of the briefs filed by the Attorney General and adopts the philosophy of the brief as its decision" (Albright 1954, 2).

## CONCLUSION

A politically sustainable judicial activism can be understood as a vehicle of regime enforcement. The idea of judicial review as regime enforcement has increasingly been developed in the literature in the context of "judicial entrenchment," or the continued enforcement by an electorally insulated judiciary of the constitutional and policy commitments of a dominant political coalition against new political majorities after the original coalition has suffered electoral defeat (Gillman 2002; Hirschl 2004). From a narrow Dahlian perspective, the active exercise of judicial review is evidence of an unruly Court hostile to the interests of the lawmakers currently in power. The "obstruction" of electoral defeat provides the most obvious context in which a political coalition might find its ability to exert its will frustrated and therefore might turn to the courts as an alternative policymaking venue. At least in the American context, however, there are other obstructions to policy hegemony as well. Political leaders may find their ability to define the policy status quo limited well before electoral defeat. In a fractured political environment such as that of the United States, national political leaders will have incentives to support the exercise of judicial review by an ideologically sympathetic judiciary even while those political leaders are still in power. The actions of a "collaborative" Court might converge with the interests of current political leaders (Tushnet Forthcoming). Most notably, the autonomy of state governments in a federal system, entrenched interests, and fragmented political coalitions may all lead political leaders to invite and/or benefit from judicial activism that can overcome such political obstructions and enforce central ideological commitments against recalcitrant officials.

Judicial review disrupts the policy status quo. The standard assumption within normative constitutional theory and a great deal of empirical literature that the

"countermajoritarian" exercise of judicial review will be viewed with disfavor by current political leaders assumes that the status quo being disrupted reflects the policy preferences of those leaders and thus that the Court is acting in a fashion that is hostile to current majorities. There are instances of judicial review in which this assumption is clearly justified. The Supreme Court's repudiation of the early New Deal is a classic example and the very exemplar of Dahl's (1957) obstructionist, "lagging" Court.

There are other episodes of judicial review that do not fit this model and do not occur in such a context. Fragmented institutions limit the hegemony of governing coalitions, and as a result limit the ability of political leaders to insure by political means that the status quo reflects their preferences. Some governmental units may be relatively autonomous and capable of setting policy that conflicts with the preferences of such coalitions. A political system with many veto points may insulate policies from electoral change, hampering the ability of current political leaders to bring policy into line. Governing coalitions suffer from a lack of ideological purity, and as a result limit the ability of coalition leaders to act politically on all the policy preferences held by important elements of its membership. Some pivotal legislators or voting blocs may have to be accommodated even at the price of policy priorities or party principles. Momentary electoral pressures may overwhelm longer term ideological commitments, leading elected officials to "shirk" their principles in order to retain office. An ideologically friendly judiciary insulated from such competing pressures may be willing and able to act where elected officials temporize. In doing so, judges may well earn plaudits, or at least deference, from the political leaders whose hands were otherwise tied. Over the course of its history, the U.S. Supreme Court has won political support for judicial review not by acting against current governing coalitions but by working within those coalitions.

Political scientists have been skeptical of the significance of truly countermajoritarian judicial review, which would seem unlikely to find political support in a democratic political system. The "friendly" judicial activism described here may be politically sustainable in ways that classical countermajoritarian judicial activism is not. Unlike countermajoritarian judicial review, friendly judicial review would not necessarily be subverted through a political appointments process that creates a sympathetic bench, nor would it necessarily be subject to the myriad legislative instruments available to sanction a wayward Court. Indeed, such political instruments for influencing the Court may be employed so as to build or strengthen a friendly Court and make judicial review more, rather than less, likely as the regime wears on. Stymied by a gridlocked Congress, for example, the Reagan administration laid plans for making jurisprudential gains through the courts (Johnsen 2003). Its plans came partially to fruition a decade later when, for example, a more conservative Court set down limits on the powers of Congress to achieve liberal aims through federal action (Keck 2004; Whittington 2001c).

Whereas a Supreme Court that flies in the face of powerful supermajorities may well find its wings clipped, a Court that acts in implicit concert with sympathetic party or factional leaders may be protected from legislative sanction by the very veto points that make judicial review useful to a political coalition in the first place (Whittington 2003). Indeed, such a Court provides incentives to elected officials to seek to build the kind of diffuse support for the Court in the general public that public opinion scholars have emphasized as important to judicial legitimacy. It has been suggested that the Court's authority to interpret the Constitution may be particularly vulnerable when faced with what Stephen Skowronek (1993) has called a "reconstructive president," a president with expansive political authority dealing with an electorally lagging Court (Whittington 2001b). If so, then the Court's authority may be at its peak when it is operating in partnership with Skowronek's "affiliated" leader, who must manage an established but fractious political coalition while advancing the contested ideological commitments of the political regime. An enterprising Supreme Court may be able to "interpose its friendly hand" to assist the political task of such an affiliated leader while exercising its independent power of judicial review.

## REFERENCES

- Ackerman, Bruce. 1991. *We the People. Foundations*, vol. 1. Cambridge: Harvard University Press.
- Adams, Charles Francis, Jr. 1875. "The Granger Movement." *North American Review* 120 (April): 394-423.
- Albright, Robert C. 1954. "Southerners Assail High Court Ruling." *Washington Post*, 18 May, 2.
- Albright, Robert C. 1956. "Rest of Party Might Just Write Off the South." *Washington Post*, 1 April, E1.
- Andrews, Edmund L. 1995. "Senate Supports Severe Penalties on Computer Smut." *The New York Times*, 15 June, A1.
- Anonymous. 1890. "Notes of Recent Decisions: Regulation of Railway Fares." *American Law Review* 24 (May-June): 516-18.
- Arnold, R. Douglas. 1990. *The Logic of Congressional Action*. New Haven: Yale University Press.
- Baker v. Carr. 1962. 369 U.S. 186.
- Banning, Lance. 1995. *The Sacred Fire of Liberty: James Madison and the Founding of the Federal Republic*. Ithaca: Cornell University Press.
- Barrow, Deborah J., Gary Zuk, and Gregory Gryska. 1996. *The Federal Judiciary and Institutional Change*. Ann Arbor: University of Michigan Press.
- Bensel, Richard F. 2000. *The Political Economy of American Industrialization, 1877-1900*. New York: Cambridge University Press.
- Berman, William. 1970. *The Politics of Civil Rights in the Truman Administration*. Columbus: Ohio State University Press.
- Bryant, Howard, and David Plotnikoff. 1996. "How the Decency Fight Was Won." *San Jose Mercury News*, 3 March, 1D.
- Burns, James MacGregor. 1963. *The Deadlock of Democracy: Four-Party Politics in America*. Englewood Cliffs, NJ: Prentice-Hall.
- Bush, George W. 2002. Statement on Signing the Bipartisan Campaign Reform Act of 2002. *Weekly Compilation of Presidential Documents*, vol. 38, no. 13, p. 517.
- Caldeira, Gregory A. 1986. "Neither the Purse Nor the Sword: Dynamics of Public Confidence in the Supreme Court." *American Political Science Review* 80 (December): 1209-26.
- Caldeira, Gregory A., and Donald J. McCrone. 1982. "Of Time and Judicial Activism." In *Supreme Court Activism and Restraint*, eds. Stephen Halpern and Charles Lamb. Lexington, MA: Lexington Books.
- Casper, Jonathan D. 1976. "The Supreme Court and National Policy Making." *American Political Science Review* 70 (March): 50-63.

- Cleveland, Grover. 1933. *Letters of Grover Cleveland*, ed. Allan Nevins. Boston: Houghton Mifflin Company.
- Clinton, William J. 1997. *Public Papers of the Presidents of the United States: William J. Clinton, 1996*. Washington, DC: Government Printing Office.
- Clinton, William J. 1998. *Public Papers of the Presidents of the United States: William J. Clinton, 1997*. Washington, DC: Government Printing Office.
- Colegrove v. Green, 328 U.S. 549 (1946).
- Congressional Record. 1894. 53rd Cong., 2nd sess., vol. 26, pt. 7.
- Congressional Record. 1995. 104th Cong., 1st sess., vol. 141, pt. 3.
- Cortner, Richard C. 1970. *The Apportionment Cases*. Knoxville: University of Tennessee Press.
- Cox, Gary W., and Mathew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.
- Dahl, Robert. 1957. "Decision-Making in a Democracy: The Supreme Court as a National Policy-Maker." *Journal of Public Law* 6 (Fall): 284-95.
- De Figueiredo, John M., and Emerson Tiller. 1996. "Congressional Control of the Courts: A Theoretical and Empirical Analysis of Expansion of the Federal Judiciary." *Journal of Law and Economics* 39 (October): 435-62.
- Delton, Jennifer A. 2002. *Making Minnesota Liberal: Civil Rights and the Transformation of the Democratic Party*. Minneapolis: University of Minnesota Press.
- Deudney, Daniel. 1995. "The Philadelphia System: Sovereignty, Arms Control, and Balance of Powers in the American States-Union, Circa 1787-1861." *International Organization* 49 (Spring): 191-228.
- Dillon, John F., ed. 1903. *John Marshall: Life, Character, and Judicial Services as Portrayed in the Centenary and Memorial Addresses and Proceedings throughout the United States on Marshall Day*. 2 vols. Chicago: Callaghan and Company.
- Dyer v. Abe. 1956. 138 F. Supp. 220 (D. Hawaii).
- Eggert, Gerald G. 1974. *Richard Olney: Evolution of a Statesman*. University Park: Pennsylvania State University Press.
- Elster, Jon. 2000. *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. New York: Cambridge University Press.
- Employers' Liability Cases. 1907. 207 U.S. 463.
- Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.
- Ferejohn, John. 1999. "Independent Judges, Dependent Judiciary: Explaining Judicial Independence." *Southern California Law Review* 72 (January/March): 353-84.
- Field, Stephen J. 1890. "The Centenary of the Supreme Court of the United States." *American Law Review* 24 (May-June): 351-68.
- Fiorina, Morris P. 1986. "Legislator Uncertainty, Legislative Control, and the Delegation of Legislative Power." *Journal of Law, Economics, and Organization* 2 (Spring): 33-51.
- Fleming, Roscoe. 1959. "America's Rotten Boroughs." *Nation*, 10 January, 26-27.
- Frymer, Paul. 1999. *Uneasy Alliances: Race and Party Competition in America*. Princeton: Princeton University Press.
- Frymer Paul. 2003. "Acting When Elected Officials Won't: Federal Courts and Civil Rights Enforcement in U.S. Labor Unions, 1935-85." *American Political Science Review* 97 (August): 483-99.
- Gates, John B. 1987. "Partisan Realignment, Unconstitutional State Politics, and the U.S. Supreme Court: 1837-1964." *American Journal of Political Science* 31 (May): 259-80.
- Gates, John B. 1992. *The Supreme Court and Partisan Realignment: A Macro- and Microlevel Perspective*. Boulder, CO: Westview Press.
- Gibbons v. Ogden. 1824. 22 U.S. 1.
- Gillman, Howard. 2002. "How Parties Can Use the Courts to Advance Their Agendas: Federal Courts in the United States, 1875-91." *American Political Science Review* 96 (September): 511-24.
- Gillon, Steven M. 1987. *Politics and Vision: The ADA and American Liberalism, 1947-85*. New York: Oxford University Press.
- Ginsburg, Tom. 2003. *Judicial Review in New Democracies: Constitutional Courts in Asian Cases*. New York: Cambridge University Press.
- Graber, Mark A. 1993. "The Nonmajoritarian Difficulty: Legislative Deference to the Judiciary." *Studies in American Political Development* 7 (Spring): 35-73.
- Graber, Mark A. 1998. "Federalist or Friends of Adams: The Marshall Court and Party Politics." *Studies in American Political Development* 12 (October): 229-66.
- Graber, Mark A. 2000. "The Jacksonian Origins of Chase Court Activism." *Journal of Supreme Court History* 25 (March): 17-39.
- Hamilton, Alexander. 1961. "Federalist No. 78." In *The Federalist Papers* by Alexander Hamilton, James Madison, and John Jay, ed. Clinton Rossiter. New York: New American Library.
- Hansen, John Mark. 1991. *Gaining Access: Congress and the Farm Lobby, 1919-81*. Chicago: University of Chicago Press.
- Hendrickson, David C. 2003. *Peace Pact: The Lost World of the American Founding*. Lawrence: University Press of Kansas.
- Hirschl, Ran. 2004. *Towards Juristocracy: The Origins and the Consequences of the New Constitutionalism*. Cambridge: Harvard University Press.
- James, Scott C. 2000. *Presidents, Parties, and the State: A Party-System Perspective on Democratic Regulatory Choice, 1884-1936*. New York: Cambridge University Press.
- Johnsen, Dawn E. 2003. "Ronald Reagan and the Rehnquist Court on Congressional Power: Presidential Influences on Constitutional Change." *Indiana Law Journal* 78 (Winter/Spring): 363-412.
- Keck, Thomas M. 2004. *The Most Activist Supreme Court in History: The Road to Modern Judicial Conservatism*. Chicago: University of Chicago Press.
- Kennedy, John F. 1958. "The Shame of the States." *New York Times Magazine*, 18 May, 12, 37-40.
- Kennedy, John F. 1963. *Public Papers of the Presidents of the United States: John F. Kennedy, 1962*. Washington, DC: Government Printing Office.
- Kennedy, John F. 1964. *Public Papers of the Presidents of the United States: John F. Kennedy, 1963*. Washington, DC: Government Printing Office.
- Kennedy, John Pendleton. 1850. *Memours of the Life of William Wirt, Attorney-General of the United States*, New and Revised Edition. 2 vols. Philadelphia: Lea and Blanchard.
- Kiewiet, D. Roderick, and Mathew D. McCubbins. 1991. *The Logic of Delegation: Congressional Parties and the Appropriations Process*. Chicago: University of Chicago Press.
- Klarman, Michael J. 1996. "Rethinking the Civil Rights and Civil Liberties Revolutions." *Virginia Law Review* 82 (February): 1-67.
- Klarman, Michael J. 1997. "Majoritarian Judicial Review: The Entrenchment Problem." *Georgetown Law Journal* 85 (February): 491-553.
- Knapp, Fred. 1997. "Exon Laments Porn Ruling Internet." *Lincoln Journal Star*, 27 June, A1.
- Krehbiel, Keith. 1991. *Information and Legislative Organization*. Ann Arbor: University of Michigan Press.
- Kutler, Stanley I. 1968. *Judicial Power and Reconstruction Politics*. Chicago: University of Chicago Press.
- Landes, William, and Richard Posner. 1975. "The Independent Judiciary in an Interest-Group Perspective." *Journal of Law and Economics* 18 (December): 875-911.
- Lewis, Anthony. 1958. "Legislative Apportionment and the Federal Courts." *Harvard Law Review* 71 (April): 1057-98.
- Lindsay, John J. 1962. "The Underprivileged Majority." *Nation*, 10 March, 208-10.
- Lovell, George I. 2003. *Legislative Deferrals: Statutory Ambiguity, Judicial Power, and American Democracy*. New York: Cambridge University Press.
- Madison, James. 1910. *The Writings of James Madison*, ed. Gaillard Hunt. Vol. 9. New York: G.P. Putnam's Sons.
- Madison, James. 1961. "Federalist No. 39." In *The Federalist Papers*, by Alexander Hamilton, James Madison, and John Jay, ed. Clinton Rossiter. New York: New American Library.
- Marshall, Burke. 1964. *Federalism and Civil Rights*. New York: Columbia University Press.
- Martin, John Barlow. 1977. *Adlai Stevenson and the World: The Life of Adlai E. Stevenson*. Garden City, NY: Doubleday.
- Martin, John Frederick. 1979. *Civil Rights and the Crisis of Liberalism: The Democratic Party, 1945-1976*. Boulder, CO: Westview Press.
- Matusow, Allen J. 1984. *The Unraveling of America: A History of Liberalism in the 1960s*. New York: Harper and Row.
- Maxfield, Sylvia. 1997. *Gatekeepers of Growth: The International Political Economy of Central Banking in Developing Countries*. Princeton: Princeton University Press.
- McConnell v. Federal Election Commission. 2003. 540 U.S. 93.
- McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28 (February): 165-79.

- McMahon, Kevin J. 2003. *Reconsidering Roosevelt on Race: How the Presidency Paved the Road to Brown*. Chicago: University of Chicago Press.
- Melnick, R. Shep. 1994. *Between the Lines. Interpreting Welfare Rights*. Washington, DC: Brookings Institution.
- Moravcsik, Andrew. 2000. "The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe." *International Organization* 54 (Spring): 217-52.
- Moustafa, Tamir. 2003. "The Judicialization of Politics in Egypt." *Law and Social Inquiry* 28 (Fall): 883-930.
- Nagel, Stuart S. 1965. "Court-Curbing Periods in American History." *Vanderbilt Law Review* 18 (June): 925-44.
- Neuberger, Richard L. 1941. "Our Gerrymandered States." *Nation*, 1 February, 127-28.
- Pickens, J. Mitchell. 2004. *Constitutional Deliberation in Congress: The Impact of Judicial Review in a Separated System*. Durham, NC: Duke University Press.
- Pickens, J. Mitchell, and Cornell W. Clayton. 2004. "The Rehnquist Court and the Political Dynamics of Federalism." *Perspectives on Politics* 2 (June): 233-48.
- Pollack v. Farmers' Loan and Trust. 1895a. 157 U.S. 429.
- Pollack v. Farmers' Loan and Trust. 1895b. 158 U.S. 601.
- Powe, Lucas A., Jr. 2000. *The Warren Court and American Politics*. Cambridge: Harvard University Press.
- Powers, Fred Perry. 1890. "Recent Centralizing Tendencies in the Supreme Court." *Political Science Quarterly* 5 (September): 389-410.
- Rakove, Jack N. 1996. *Original Meanings. Politics and Ideas in the Making of the Constitution*. New York: Vintage Press.
- Ramseyer, J. Mark. 1994. "The Puzzling (In)dependence of Courts: A Comparative Approach." *Journal of Legal Studies* 23 (June): 721-47.
- Reno v. American Civil Liberties Union. 1997. 521 U.S. 844.
- Reston, James. 1952. "Governor Says 'No'." *New York Times*, 31 March, 1.
- Reynolds v. Sims. 1964. 377 U.S. 533.
- Richardson, James D. 1908. *A Compilation of the Messages and Papers of the Presidents, 1789-1907*. 11 vols. New York: Bureau of National Literature and Art.
- Rogers, James R. 2001. "Information and Judicial Review: A Signaling Game of Legislative-Executive Interaction." *American Journal of Political Science* 45 (January): 84-99.
- Rosenberg, Gerald N. 1991. *The Hollow Hope: Can Courts Bring About Social Change*. Chicago: University of Chicago Press.
- Rosenberg, Gerald N. 1992. "Judicial Independence and the Reality of Political Power." *Review of Politics* 54 (Summer): 369-98.
- Rowe, James, Jr. 1995. "The Politics of 1948." In *Documentary History of the Truman Administration*, edited by Dennis K. Merrill. Vol. 14. Bethesda, MD: University Publications of America.
- Schattschneider, E.E. 1975. *The Semi-Sovereign People: A Realist's View of Democracy in America*. Hinsdale, IL: Dryden Press.
- Schwartz, Bernard. 1983. *Super Chief: Earl Warren and His Supreme Court—A Judicial Biography*. New York: New York University Press.
- Schwartz, John. 1996. "Language on 'Indecency' Sparks Telecommunications Bill Protest." *Washington Post*, 5 February, A8.
- Shapiro, Martin. 1999. "The Success of Judicial Review." In *Constitutional Dialogues in Comparative Perspective*, ed. Sally J. Kenney, William M. Reisinger, and John C. Reitz. New York: St. Martin's Press.
- Skowronek, Stephen. 1993. *The Politics Presidents Make: Leadership from John Adams to George Bush*. Cambridge: Harvard University Press.
- Sowell, Randy L. 1992. "Judicial Vigor. The Warren Court and the Kennedy Administration." Ph.D. diss. University of Kansas.
- Special to *The New York Times*. 1954. "Tennessee Race is Growing Heated." *The New York Times*, 25 July, 60.
- Special to *The New York Times*. 1960. "U.S. Asks High Court Review of Tennessee Apportionment." *The New York Times*, 20 April, 25.
- Special to *The New York Times*. 1962. "Capital is Split on Apportioning." *The New York Times*, 28 March, 1.
- Spriggs, James F., II, and Paul J. Wahlbeck. 1995. "Calling It Quits: Strategic Retirement on the Federal Courts of Appeal, 1893-1991." *Political Research Quarterly* 48 (September): 573-97.
- Stephenson, Donald Grier, Jr. 1999. *Campaigns and the Court: The U.S. Supreme Court in Presidential Elections*. New York: Columbia University Press.
- Stimson, James A., Michael B. Mackuen, and Robert S. Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89 (September): 543-65.
- Summers, Festus P. 1953. *William L. Wilson and Tariff Reform. A Biography*. New Brunswick, NJ: Rutgers University Press.
- Sundquist, James L. 1983. *Dynamics of the Party System: Alignment and Realignment of Political Parties in the United States*, Revised Edition. Washington, DC: Brookings Institution.
- The New York Times*. 1894. "Mr. Cockran's Final Effort," 31 January, 6.
- The New York Times*. 1895a. "Psychology of the Supreme Court," 2 May, 4.
- The New York Times*. 1895b. "Democratic Doctrine Destroys the Populist Income Tax," 21 May, 4.
- The New York Times*. 1896. "Clevelandism Again," 21 November, 4.
- Omaha World Herald*. 1997. "Exon's Decency Act a Good Try, But Sleaze Difficult to Fight," 29 June, 12b.
- Tiedeman, Christopher G. 1886. *A Treatise on the Limitations of Police Powers in the United States*. St. Louis: F.H. Thomas Law Book Co.
- Truman, Harry S. 1964. *Public Papers of the Presidents of the United States 1948*. Washington, DC: Government Printing Office.
- Truman, Harry S. 1968. *Public Papers of the Presidents of the United States: 1952-1953*. Washington, DC: Government Printing Office.
- Tushnet, Mark. Forthcoming. "The Supreme Court and the National Political Order: Collaboration and Confrontation." In *The Supreme Court in American Political Development*, ed. Ronald Kahn and Ken I. Kersch. Lawrence: University Press of Kansas.
- Vanberg, Georg. 2001. "Legislative-Judicial Relations: A Game Theoretic Approach to Constitutional Review." *American Journal of Political Science* 45 (April): 346-61.
- Voigt, Stefan, and Eli M. Salzberger. 2002. "Choosing Not to Choose: When Politicians Choose to Delegate Powers." *Kyklos* 55 (2): 289-310.
- Warren, Charles. 1926. *The Supreme Court in United States History*. 2 vols. Boston: Little, Brown, and Company.
- Warren, Earl. 1977. *The Memoirs of Earl Warren*. Garden City, NY: Doubleday.
- Weaver, R. Kent. 1986. "The Politics of Blame Avoidance." *Journal of Public Policy* 6 (4): 371-98.
- Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91 (June): 245-63.
- Welsh, Orville. 1932. "The Democratic Revolution." *Nation*, 30 November, 523-24.
- Weisberg, Jacob. 1996. "Strange Webfellows." *Slate*, 14 December (<http://slate.msn.com/id/2257>).
- White, Walter. 1948. *A Man Called White. The Autobiography of Walter White*. New York: Viking.
- Whittington, Keith E. 2001a. "The Road Not Taken: *Dred Scott*, Judicial Authority, and Political Questions." *Journal of Politics* 63 (May): 365-91.
- Whittington, Keith E. 2001b. "Presidential Challenges to Judicial Supremacy and the Politics of Constitutional Meaning." *Polity* 33 (Spring): 365-95.
- Whittington, Keith E. 2001c. "Taking What They Give Us: Explaining the Court's Federalism Offensive." *Duke Law Journal* 51 (October): 477-520.
- Whittington, Keith E. 2003. "Legislative Sanctions and the Strategic Environment of Judicial Review." *International Journal of Constitutional Law* 1 (July): 446-74.
- Whittington, Keith E., and Daniel P. Carpenter. 2003. "Executive Power in American Institutional Development." *Perspectives on Politics* 1 (September): 495-513.
- Whittington, Keith E. Forthcoming. "Preserving the 'Dignity and Influence of the Court': Political Supports for Judicial Review in the United States." In *The Art of the State: Rethinking Political Institutions*, ed. Ian Shapiro and Stephen Skowronek. New York: New York University Press.

# Political Science on the Cusp: Recovering a Discipline's Past

JOHN G. GUNNELL *State University of New York at Albany*

*As Thomas Kuhn noted, it is almost inevitable that scientific practitioners read the history of their field backward and perceive earlier stages as, at best, prototypical of the present. This is the manner in which political scientists, and even historians, have imaged the relationship between the debates about science and democracy that took place during the 1920s and 1950s. Despite the importance of Charles Merriam's role in the history of American political science, his work was not the discursive axis of the paradigmatic disciplinary shift that took place in the first quarter of the 20th century. It was the arguments of G. E. G. Catlin and W. Y. Elliott that most distinctly represented the transformation in both the theory of democracy and the image of science, and that, for the next two generations, set the terms of the debate about these issues as well as about the relationship between the mainstream discipline and the subfield of political theory. And, despite the theoretical and ideological differences between Catlin and Elliott, their exchange points to the intensely practical concerns that originally informed the controversy about the scientific study of politics.*

The theory of democracy and the conception of scientific inquiry that are often assumed to have emerged in the course of the behavioral movement, during the 1950s and 1960s, were not only initiated in the 1920s but also far more evolved than typically acknowledged by contemporary political scientists and in historical scholarship. Not only has the significance of the disciplinary transformation that gave rise to these earlier developments been overlooked, but also the pivotal conversation, represented by the work of G. E. G. Catlin and William Yandell Elliott, has been obscured. Many dimensions of contemporary political science and debates revolving around such issues as the nature of scientific explanation, the role of political theory, the concept of democracy, and the relationship between political science and politics cannot be adequately understood without a more accurate etiology.

By the end of the first quarter of the twentieth century, the discipline and profession of political science were, as one major participant in discussions about the condition of the field put it, at a "crossroads" (Ellis 1927). Such an assertion may not be uncommon, but after more than three-fourths of a century, it remains an astute diagnosis of the period. Ellen Deborah Ellis was the first person to articulate the exact character of the theoretical crisis that had been emerging in the discipline, but, in the same year, Charles Beard, in his Presidential Address, argued that creativity in political science had been stifled by the conservative forces of legal and historical studies and by the pressures of professionalization and specialization (Beard 1927a). Although this period has been recognized as representing an important intellectual intersection in the history of the field, the general coordinates of that junction and the contours of the road chosen have not been adequately examined and reconstructed. This era has

often been characterized, by both political scientists and historians of social science, as marked by a protobehavioral revolution involving, whether assessed as pejoration or progress, a turn away from institutional studies and the methods of history and toward science and quantitative analysis (e.g., Almond 1996; Crick 1959; Ricci 1984; Seidelman 1985; Ross 1991; Somit and Tanenhaus 1967). Although such accounts have identified important attributes of this transition in the theory and practice of political science, they have often been advanced in the context of judgments and debates about the present state of the discipline and served as vehicles of legitimation and critique (Dryzek and Leonard 1988). The retrospective imposition of contemporary categories has inhibited identifying the conceptual core of the transformation and assessing its implications for understanding, and critically reflecting on, the present character of the field.

If there has been any point in the history of political science that can be construed as what Thomas Kuhn (1970) referred to as a paradigmatic shift, it was the 1920s. Although there has been a great deal of ambiguity and controversy about what Kuhn meant by the term "paradigm" and about how, and the extent to which, his analysis could be applied to the history of the social sciences, the most salient characteristics of his account of a scientific revolution were manifest in the discipline of political science during this period, and arguably *only* in this period (Gunnell 2004a). What in part explains the failure of political scientists to understand this phase adequately is what Kuhn (1970) referred to as the "invisibility" of revolutions as they become sublimated in the typical practitioners' image of their enterprise as a linear, cumulative, and progressive endeavor. It is this disposition which "systematically disguises the existence and significance of scientific revolutions." This internal perception and the "authority" attaching to it has often been incorporated in more external historical narratives, and at all levels, "the temptation to write history backward is both omnipresent and perennial" (Kuhn 1970, 136, 138). Although what has been conventionally self-ascribed and designated as the behavioral revolution was innovative in many respects and had a deep and lasting impact on

John G. Gunnell is Professor, Department of Political Science, State University of New York, Albany, NY 12222 (jgg@albany.edu).

An early version of this essay was presented at the 2004 meeting of the Western Political Science Association. The final version has benefited significantly from the suggestions and constructive criticism of the editor and three anonymous reviewers as well as from conversations with David Easton and David Elliott.

what Kuhn referred to as the “disciplinary matrix,” it was more a “reformation” than a revolution. It was to a significant degree a response to a new and unprecedented wave of critical scholarship that challenged the prevailing paradigm, but it served, intellectually and professionally, to solidify, vouchsafe, and institutionalize an earlier and more fundamental change in both the conception of political reality and the theory of democracy that, by the early 1930s, had structured the field (Gunnell 2004b, 1993). The behavioral movement was, in effect, an attempt to complete the transition of the discipline to a practice of “normal science.”

According to Kuhn (1970, 5, 94, 110–11), the essence of a scientific revolution is a theoretical reconstitution of the phenomena that are the object of investigation—“when paradigms change, the world itself changes with them,” because the activity of science “is predicated on the assumption that the scientific community knows what the world is like.” Kuhn argued that “a scientific theory is declared invalid only if an alternative is available to take its place” (77), and, as he stressed in his later work, the core of his argument turned less on the meaning of “paradigm,” and whether it referred to a general theory or to concrete “exemplars,” than on his claim about the fundamental “incommensurability” of the “kind-concepts” that define the “lexical structure” of a scientific “language community” which is “constitutive of *possible experience* of the world” and the practice of science (Kuhn 1993, 328–33; 2000). It was just such a theoretically grounded shift among language communities that occurred in political science during the 1920s.

The nineteenth century theory of the state, as articulated by theorists such as Francis Lieber, Theodore Woolsey, and John W. Burgess, was both a general theory of politics and a theory of democracy, and it was the basis of an elaborate descriptive and explanatory account of the history and practice of both American politics and political science. This theory, which had in large measure defined the discipline and profession since its inception, assumed that democracy required the existence of a relatively homogeneous “people” as the author and subject of self-government, and it was this concept of political community for which these theorists had reserved the term “State,” from which they pointedly distinguished the institutions of “government.” By the end of the first decade of the twentieth century, this theory had all but dissolved, and this precipitated a “crisis” as such factors as an increasing recognition of social heterogeneity and a growing identification of “state” and “government” appeared as “anomalies.” The theory of the state, however, did not simply fade away and leave a theoretical and methodological vacuum. Despite a rejection of its philosophical and ideological foundations, the remnants of this theory persisted among Progressive political scientists and philosophers such as John Dewey (1927b), who continued to maintain that democracy was only viable if modern society, with all its diversity and complexity, was transformed into a self-conscious “public” or “great community.” Such a claim, however, was what pluralists such as John Dickinson (1930) referred to

as the “democratic dogma” that must be replaced by a vision of “democratic realities.” These “realities” were based on an empirical pluralist account of politics joined to a normative theory of democratic pluralism. The pluralist attack on the idea of state sovereignty was much more than a rejection of legalistic and institutionalist analysis. It was, in effect and as well, an attack on the prevailing image of *popular* sovereignty. Although the discipline had always claimed the status of science, the criteria of what constituted scientific claims and appropriate methodology became contested in this confrontation between theories, and these issues were integrally related to what had become the pressing problems of disciplinary identity and the practical as well as the cognitive relationship between political science and politics.

In the case of any scientific revolution, there is, as Kuhn (1970) noted, usually some conversation that is at the center of the shift and associated with particular individuals, but political scientists have often looked in the wrong place for this discursive axis. The belief that the work of Charles Merriam most fundamentally represented the transformation in American political science that took place in the first quarter of the twentieth century has become a dominant piece of academic folklore, but although the ascription of importance to Merriam’s work is certainly merited, particularly with respect to his role as the impresario of the new science of politics, it was not the locus of the theoretical turn. Merriam was resolute in never characterizing his predecessors as representing anything other than a stage in the cumulative progress of inquiry, and despite his ideological divergence from his mentors (such as Burgess and William Archibald Dunning), he not only refrained from criticizing their work but also praised it as a major scientific advance. Furthermore, as in the case of many Progressives, his image of democracy and emphasis on such things as civic education remained closer to the nineteenth century *Staatslehre* than to the emerging pluralist paradigm. Merriam’s account of scientific explanation was conceptually thin and had little to do with the philosophy of science and image of theory construction that would eventually dominate the field and be represented in the deployment and justification of approaches to inquiry such as systems analysis and rational choice theory. Despite his commitment to scientism, Merriam never, for example, depreciated the study of the history of political theory, and it was not his most prominent critics who established the form of discourse that would persist, through the 1960s, as the intellectual counterpoint to the emerging ideas of both democracy and science. The question, then, is that of where to seek the principal interlocutors in the disciplinary revolution that had taken place by the late 1920s and which, for the remainder of the century, would structure the evolution of both the normative and the empirical dimensions of mainstream political science.

There were two remarkable, but now largely forgotten, or at least neglected, books in which the fundamental issues were most clearly joined and crystallized: Catlin’s (1927c) *The Science and Method of Politics*



and Elliott's (1928b), *The Pragmatic Revolt in Politics*. These works exercised exemplary and determinative roles in the disciplinary conversations of this period, and the exchanges between the authors uniquely manifested the theoretical and ideological tensions within the field, the search for the autonomy of political science among the social sciences, and concerns about the relationship between political science and public life. They also represented a period of intense Americanization and Anglicanization, after the discipline had, for nearly a century, been beholden to German philosophy. Whereas Elliott was an American who received his D.Phil. at Oxford, Catlin was an Englishman who took his Ph.D. at Cornell. In England, Elliott found an image of politics and political studies that, with emendations, he imported to the United States in support of the reconstitution of a waning image of democracy and political science. Catlin, while holding on to motives and motifs formed in Britain, escaped from what he perceived as the limitations of English scholarship in order to seek in America the grail of a more scientific approach to the study of politics that would serve to validate and realize a new vision of popular government. Their books were revisions of their dissertations, and both individuals, during the entire course of their careers, were deeply involved in practical politics.

## CROSSING OVER

In order to serve in France during the World War I, Catlin, in 1914, vacated an Oxford "exhibition," once held by Harold Laski. While eventually waiting "recall" to Oxford and critically reflecting on what he took to be the limitations of what he had learned at that institution, he began to formulate the ideas that would eventually inform his book. He later remarked that, during the war, "at least we lived in a real world, unlike that of some of the Oxford society which for me was to follow." His examiner at Oxford was Ernest Barker, who had demoted Catlin's scholarship to an "exhibition," because, Catlin claimed, "he thought I plagiarized his own ideas." Although he was offered fellowships at both Exeter and the University of Minnesota, he eventually took a position at Sheffield, but when informed that the college did not, and would not, "teach political philosophy and science," he decided finally that in England "there was no room to take wing." Cornell trumped Harvard's offer of a fellowship by awarding him a "professorship" and allowing him, at the same time, to write his dissertation. Oxford, even after replacing the traditional study of the Greats, "did not display active interest" in inviting him to return, and he settled down at Cornell to focus on what he considered to be the pressing problems of world peace, the exercise of power in politics, and the nature and role of a science of politics.

Catlin took the position at Cornell in 1924, and, in the same year, with the encouragement of the Cornell psychologist Bradford Titchener, finished his dissertation. A short time later, he was invited by the Rockefeller Foundation and the Social Science Research Council

to come to New York City and study Prohibition, and it was here that he met Merriam and Harold Lasswell and subsequently not only became deeply involved with American political scientists who were dedicated to propagating a new science of politics but also gave new depth to that vision and joined it conceptually to the emerging empirical and normative pluralist paradigm. What attracted Catlin to the American scene was not simply the growing emphasis on empiricism but the manner in which this perspective was linked to a commitment to practical application which he believed had been so lacking at Oxford. With respect to the latter, Catlin found in the United States the intellectual ambience that was conducive to this goal. He considered Merriam the "Moses" of the new political science, and Arthur Bentley its "Grandma Moses," and with Lasswell, he "maintained an unbroken partnership . . . We found political science a chaos. We left it tidied up," and "with the pluralists we deposed the sovereign national state." He later took umbrage, however, at Bernard Crick's (1959) designation of the discipline as the *American Science of Politics*, because, Catlin noted, his own initial "thoughts were shaped at Oxford" by his work on Hobbes as well as through contact with individuals such as Graham Wallas and Laski. He left Cornell in 1935, but continued to pursue, until the end of his life, the connections between political science and political action (Catlin 1972).

Catlin's book was the product of a gestation period of several years before he was finally impelled to "adventure to America in order to have free time for study" (Catlin 1927c, x-xii), and it incorporated some of his early articles (Catlin 1925, 1927a, 1927b). He conceived the work as a contribution to a genre that had been "almost untouched" since Aristotle, that is, a conception of a rigorous science of politics dedicated to an end in action. Although he deemed it an "unsystematic" exercise, it was more thorough and specific than anything produced at that time by his American heroes who have been more commonly associated with the emerging scientific persuasion.

The political and intellectual context of the book, as Catlin perceived it, was the postwar era and a general "revulsion against the doctrine of the omnipotent State" as it was manifest both in politics and in the study of politics. This turn away from the theory of the state, Catlin claimed, had its roots (as Barker had already stressed) in the English Whig tradition of natural rights, tractarian religion, and, the more recent, forces of trade unionism and internationalism. These positions were reinforced by a variety of pluralist arguments which, even prior to the recent "Fascist reaction," had "confronted the theory of the absolute State power with a realistic presentation of the facts of the governmental process" and a new vision of democracy (1927c, ix). Catlin's only quarrel with pluralism (and he had Laski in mind) was that much of it was still attached to and grounded in a liberal "ethical philosophy" rather than a descriptive, empirical, and "dispassionate study of actual human behavior" and "the rules of political conduct." The task that Catlin set for himself was to find a scientific and empirical basis for



democratic theory, a search that was to define much of mainstream political science through the remainder of the century (e.g., Dahl 1956; Herring 1940). Although he believed that pluralism was “ethically superior,” it was too often approached as another abstract “ideal” which became a form of “dogmatism” and which, if it lacked scientific grounding, had “no better guarantee of issuing in practical political action” informed by a grasp of the “basal principles of political method.” Knowledge of such principles, as in the case of all “the great conquests of man over nature,” would, he argued, demonstrate what was “feasible” and provide a means of social “control” (1927c, x–xi). Although Laski had not supported his return to England, Catlin noted that he was indebted to Laski as well as to Barker but also to the American sociologist Harry Elmer Barnes who had been one of the early proponents of both normative and descriptive theories of pluralism as well as the introduction of empirical analysis into political theory.

Although it is widely, but incorrectly, believed that, during the 1920s, political scientists explicitly rejected historical studies in favor of the pursuit of a more rigorous science, Catlin was the only person to undertake an extensive critique of history in these terms. Although historical research, he argued, should properly seek to find out what had happened in the past, as a prelude to gaining control of life in the present, it had been bogged down in an unmethodical collection of particular “facts” that revealed little of the “real world” in any manner that could contribute to a “scientific study of politics” (7, 22). Although in the physical sciences, “data have long been arranged with a view to the yielding of practical and theoretical results,” historical study, he argued, was a “chaos” dominated by the anecdotal reflections of moralists, artists, idealists, and antiquarians. Even the “scientific” school of history mistook “accuracy” in the collection of facts for real science and failed to provide any “utilitarian treatment” (23, 55, 66). If history was to be “useful,” it was necessary to “generalize” and compare and to focus on its “contemporary,” “institutional,” and “psychological” dimensions. History, itself, Catlin argued, “is and never can be a science; but it provides the data for the social sciences” and must be “scientifically prepared” through “the deductive method with a hypothesis” and by techniques of measurement and quantification which would make it possible “to administer appropriate and scientific legislative and administrative remedies” in the pursuit of “social therapy” (1927c, 67, 70–71, 74–76, 81–82, 84).

Catlin’s principal argument was for the adoption and application of what he took to be a universal scientific method of generalizing from particulars, which could be the basis of an authentic law-governed science of politics. When, however, he employed phrases such as “the method” or “methodology of politics,” he was usually referring to the subject matter of politics and to a general “rational ‘form’ or schema” which, he argued, could be construed as informing individual and group behavior—what he would designate as the exercise of will and the search for power. What was required, he argued, much as David Easton (1953) would claim a

generation later in his critiques of “hyperfactualism” and the poverty of general theory, was a “*Theory of Social Structure*” and a “grasp of the fact that social phenomena are interconnected in a social system.” Such an approach, he insisted, would yield a useful definition of politics and a basis for “observations of social behavior” and the “study of the political structure.” Catlin claimed that this kind of theory had already been achieved by economists through their model of “economic man” and that a similar strategy should be utilized in the study of politics (1927c, 83–85).

In defending the “possibility of a political science,” Catlin drew upon Karl Pearson’s account of theoretical instrumentalism, that is, the claim that scientific theories and laws are conventional constructs and constitute a “logical structure superimposed on the observation of a highly frequent occurrence.” Theories and facts were viewed as separate, yet theories were held to be both derived from and tested by factual observations. This image of theory would be further elaborated by later positivist and logical empiricist philosophers of science, and, in secondary and tertiary form, it would become the received account of science characteristic of mainstream political science during the behavioral era (e.g., Easton 1965; Lasswell and Kaplan 1950). Catlin noted that although scientific methods were often unpopular and misunderstood, it was now important for “Politics” to “clear for itself the forest of detail by the use of abstract hypotheses and of a scientific method.” One of the things that would separate Elliott and Catlin, and subsequently antibehavioralists and behavioralists, was the latter’s commitment to the idea that neither politics nor political inquiry was substantively autonomous. Following Wallas, and Walter Lippmann, Catlin argued that it was really human nature and society that were at issue, and Catlin, like behavioralists after midcentury, took politics to be a domain that was analytically distinguished and constituted. He even claimed, similar to later political scientists such as Robert Dahl (1956), that the institution of “Government is a relatively unimportant part of Politics,” if compared with the significance of social processes such as group interaction and the pursuit of power (1927c, 91, 93, 95, 99).

Catlin maintained that because human nature was complex, a science of politics required a radically interdisciplinary approach rather than divided “departmental knowledge.” A predictive science should be modeled on the form of the natural sciences and entailed that “Politics must view social phenomena externally.” A science of “pure politics is limited to naturalism” and “consists of a body of verifiable and systematic knowledge gathered by observation and experiment” rather than “a conglomerate of historical excursus, of *belles lettres* about ‘liberty’ and the like, and of debating points prepared for a party platform.” In order to simplify the facts of observation, a concept of the political “process” was required rather than a focus on, what he considered to be, ghostly entities such as the state as it had been conceptualized in the past. What should be studied was the statistically constituted “‘averageness’ of human nature” and “the conduct of the average

man" and group organized around assumed motives such as the pursuit of self-interest. Catlin believed that such a notion of instrumental rationality could be detached from the ideology of classical economics and become the basis of a hypothetical "working fiction" of "an abstract political man" or "laboratory creature" representing a "conscious formalism." Such a model was not to be construed as a claim about reality but as a tool that had explanatory utility and served the function of undermining "uncritical theories" based on obsolete concepts such as sovereignty. This, then, was his vision of the possibility of political science, but, Catlin argued, "there is yet no such thing as a political science in any admissible sense," and "hitherto the unguarded field of political theory has been a veritable Valley of Hinnon wherein men have been permitted to cast without challenge the rubbish of uncritical speculation and the burning oil of enthusiasm, to fling the bodies of opponents and to sacrifice to strange idols" (1927c, 106-7, 112-13, 126, 130-31, 141-44).

Because, in Catlin's view, there was no indigenous subject matter of political science, which predetermined its domain, he was forced to confront the issue of the discipline's "place among the social sciences." Although he noted that many fields such as law and history had attempted to annex and submerge the "middle province" of politics, there were, he claimed, sufficient heuristic grounds for viewing it as "a realm in its own right." Although it might seem that "to pass from political to economic theory is like passing from sea-fog to mountain air," there was reason to move beyond an economic interpretation of history to a political one and to focus on political structures comparable to what Durkheim had spoken of as objective social facts. Yet, as he suggested that E. R. A. Seligman and Beard had demonstrated, in their studies of economics, politics, and history, attempts either to define "social determinants" in terms of strict disciplinary boundaries or to unify the social sciences were mistakes and bound to fail. There was "not one social science but many jostling each other," and any hope "precisely to define frontiers must be futile." He claimed, nevertheless, that because the objects of social science were analytically delimited, political science had as much standing as any other discipline and could carve out a domain based on a concept of "man in his relation to the wills of his fellows in control, submission and accommodation" (1927c, 163-64, 168, 205).

Catlin's most detailed and focused argument, which, within political science, was the principal discursive forbear of contemporary rational choice analysis, was devoted to giving an account of "the process of politics" in terms of an extended analogy with economic theory. Once again, he claimed that the success of economics resided in the manner in which it had created an "ideal being," and "whether economic man exists or ever existed is immaterial." Although a "mere fiction," it could function as an abstract "scientific hypothesis." The idea of a "political man," based on assumptions such as Hobbes's idea of human beings as creatures "seeking power," would be a kind of "scientific Frankenstein" who, rather than pursuing money, as in the case of

economic man, strived for "man-power" such as that involved in voting (1927c, 213-15, 253, 257).

Although there was no ambiguity about Catlin's commitment to the idea that the ultimate purpose of a science of politics was social transformation and control, or what he referred to as a "science of social hygiene" (Catlin 1927d, 252), he argued that this, paradoxically, required distinguishing between facts and values and separating political science and ethics. Pure science must precede application. In a pluralistic society, science could wield influence over politics only if it was grounded on the neutral authority of objective and impartial methods. Although, as in the case of Woodrow Wilson, Catlin sought a kind of rhetorical assimilation of theory and practice by referring to both the subject matter and the field of inquiry as "Politics," he insisted on the difference between science and values. Like Lippmann, Wallas, Merriam, Lasswell, and others, Catlin was committed to a science that would inform experts acting in the public realm. "The most important task of democratic education in politics is to inculcate an invincible skepticism about the range of lay political knowledge... The political situation must, then, be approached not with preaching and programmes, but in the attitude of a profession of social medicine" in which ethical theories should not intrude. "It is no more the task of a political scientist to instruct men about political values than it is the task of a teacher of sculpture to instruct his pupils in theories and ideals of artistic expression." The role was only to provide material and technique (Catlin 1927c, 288, 295, 299).

Catlin urged "moral detachment," the notion that "Politics is concerned with means; Ethics with ends," that Politics "studies what is, not what should be," and that values are grounded in aesthetic rather than scientific judgments. Despite what he believed was the contrary view manifest in much of the history of political thought as well as in contemporary studies of that history, he insisted that there is "no one value or system of values which alone can be held or known to be laudable or right." But, at the same time, he claimed that better values were grounded in better knowledge and that such knowledge was the answer to the salvation of civilization. Although he argued that science could tell us what will happen but not what we should do, he assumed that knowledge of what will happen was practically compelling. As in the case of Dewey and the pragmatists, he believed that the progress of science would ultimately both make clear the proper value choices and provide a means of implementing those choices (1927c, 299, 301, 310, 323, 325, 349).

Albert Einstein noted that Catlin was "one of the first in our time to treat systematically the question of linking theory with practice in politics" (Catlin 1972, 58). Dewey, reviewing Catlin's book along with the Harvard philosopher Ernest Hocking's *Man and the State*, noted that both stressed psychology, but although Hocking held on to an obsolescent defense of the state, "the impression left upon me by [Catlin's] book is one of wholesomeness, like a refreshing breeze blowing through a close atmosphere." He judged it to be

"brilliantly written" and "pragmatic" in purpose with a "brilliant critique of various schools of history-writing" (Dewey 1927a). Robert Crane (1927), a strong supporter of the new scientism, maintained that, despite some faults, Catlin's work had "the distinction of offering the sole constructive suggestion of recent years in the methodology of politics. A. Gordon Dewey (1927), another advocate of empirical approaches in political science, worried that the individualistic economic model might not take account adequately of the now accepted fact that "effective political activity is the activity of groups," but he praised the work as exemplifying the methodological direction that the discipline should follow at a time when it was in a "state of flux" and seeking to probe behind "the visible organs of government."

It is not surprising that what drew the most negative reaction were Catlin's advocacy of interdisciplinary studies and his depreciation of history. George H. Sabine (1928), who would later join the philosophy department at Cornell and whose work, a decade later, would dominate the study of the history of political theory, agreed that in some respects the general spirit of the book did reflect Aristotle's project, but he noted that the historian would be "ill-advised" to accept the role of fact-collector for social scientists and that, with respect to the adaptation of economics, this "cold-blooded imitation of one science by another" was inappropriate in the study of politics where, for example, prediction was more difficult. Beard (1927) went considerably further and suggested not only that economists would be surprised by the degree to which Catlin had credited them with powers of prediction but also that "if this procedure offers a correct working hypothesis, then darkness enshrouds all those who labor under the impression that politics is fundamentally concerned with the state and government and with the social and economic factors that appear to determine, or at all events condition, their forms and operation." What is interesting, however, is that neither Sabine nor Beard was theoretically and philosophically far removed from Catlin and that what bothered them was primarily the matter of the demarcation of academic terrain.

In 1930, Catlin published *A Study of the Principles of Politics*, which was dedicated to Titchener and Wallas and in which he more explicitly noted his debt to Dewey. This work received less attention, probably because it was in part essentially a reprise of the earlier book, but also because by this point there were few political scientists who were inclined to challenge his conceptions of either science or democracy. But he again stressed the need to turn away from typical studies of "political philosophy and the humanities" and toward the development of a naturalistic interdisciplinary science of society based on general laws, and he argued ever more strongly that such a science should be devoted to the utilitarian goal of social control through the manipulation of public opinion. Harold Gosnell (1933), one of the early innovators in the application of quantitative methods in political science, praised Catlin as a pioneer, but strangely,

one might think, Merriam (1931) had little that was positive to say about the book. Part of the problem was an underlying theoretical disagreement that centered around Laski, to whom, Merriam noted, Catlin made frequent reference and with whom he seemed concerned about "reconciling his doctrines." Merriam doubted "that what we need is another still more orthodox *Grammar of Politics*" and an attempt to reconcile "Austinians and Pluralists." This was a very odd characterization of the work, but Merriam concluded that, on the whole, the book did not demonstrate "maturity of thought, facility in generalization, penetration, and vision." What is once again important to note is that Merriam was not a defender of pluralism as an account of democracy, and that despite Catlin's insistence on distinguishing between ethical and empirical claims, he embraced pluralism as a democratic theory.

Despite Merriam's criticism, Catlin remained steadfast in his defense of the Chicago school, and he actually provided the most elaborate and articulate statement of the vision of science with which Merriam and Lasswell are commonly associated. In reviewing Lasswell's *World Politics and Personal Insecurity* Catlin noted that it was "one of the five most important books on political theory since the war" and part of a trilogy which included Merriam's *Political Power* and T. V. Smith's *Beyond Conscience*. Together these works, he suggested, signaled the triumph of one of the principal strands of political theory that, after the war, constituted the revolt against "the idealistic doctrine of the right-wing successors of Hegel." The "Chicago School of Philosophy," Catlin claimed, "became incarnate with Professors Merriam, Park, Smith and Lasswell, as the Chicago School of Politics" which, in its turn away from "political philosophies of values," and toward quantitative and interdisciplinary studies, brought to an end the contempt once directed toward the idea of a "science of politics, as a mechanics of means." Catlin predicted that this corpus would be "more important than Marxism, because it would "substitute exact, verified and impartial knowledge for political theology and messianism" and supplement principles and values with "a science of the practicable, acquainted with the limits set by human nature and natural law." In the end, he wondered only if the members of the Chicago School would not be "well advised to direct their attention to the revival of natural law, not, however, treated on the "ethical plane of the Schoolman, but on the plane of objective, scientific inquiry" (1935, 278–81). Within the next decade, the study of natural law would, with the presidency of Robert Maynard Hutchins and the appointment of individuals such as Leo Strauss, indeed, come to Chicago, but not in the manner that Catlin anticipated. In the 1920s, however, it was from Harvard that the basic critique of pluralism and the vision of a new science emanated.

## HOLDING THE LINE

In some respects, one might equate Elliott's response to the new theory of democracy, and the methodological claims attending it, to that of the eminent

Swiss/American natural scientist Louis Agassiz who contributed so much to modern science but, in the face of an overwhelming acceptance of Darwin's theory, remained one of the last holdouts for a vision of intelligent creation and its social as well as natural implications. Even Darwin, as he finished the *Origin of the Species*, did not expect any quick agreement from many of his contemporaries who had for so long conceived of the world differently, and Elliott spoke for those who could not conceive of democracy as based on other than communal unity and of political science as detached from values.

Elliott was born in Murfreesboro, Tennessee, attended private school, and graduated from Vanderbilt where, before becoming a political scientist, he was one of the Fugitive Poets. The latter were originally several friends, including, most notably, John Crowe Ransom, Allen Tate, and Donald Davidson, who met in a salon atmosphere in Nashville. The group disbanded at the beginning of World War I but came together again in 1919 as the individuals drifted back to Vanderbilt. Elliott had been associated with this gathering as early as 1915, but, along with Robert Penn Warren, joined the now more formal association shortly after they began to publish, in 1922, the poetry journal, *The Fugitive*. While studying in Königsberg, Germany, Elliott submitted a nostalgic contribution, "A Critique of Pure Reason," in which he lamented his current situation as a "metaphysicked fool" who longed for the countryside and communal culture of rural Tennessee (Elliott 1928a). The "fugitives" were more diverse than the later "Agrarians" who, after 1928, included Ransom, Davidson, Tate, and Warren and who were the authors of the manifesto *I'll Take My Stand*. The agrarians were distinctly romantic idealists seeking to recover a lost sense of community which they associated with the old South, but both groups were, on the whole, characterized by an antimodernist spirit which was reflected in Elliott's work.

Like Catlin, Elliott served in France during the war, and, like Tate and Warren, attended Oxford as a Rhodes Scholar where he eventually took his degree in the new Politics, Philosophy, and Economics curriculum initiated by A. D. Lindsay at Balliol. After teaching briefly at Berkeley, Elliott joined the Harvard faculty, at the invitation of President A. Lawrence Lowell, in 1923, at the age of 27, where he taught for 41 years before accepting a position, prior to retirement, at American University. At Harvard, he directed more than one hundred dissertations; and, along with Carl Friedrich, dominated the department, and particularly the field of political theory, for many years. Among his students were some of the most influential, but eventually diversely inclined, future members of the profession including Easton, Sheldon Wolin, Samuel Huntington, and William Riker, and he was the teacher and mentor of Henry Kissinger and Pierre Trudeau. His students often complained of his absence when, from the 1930s to the 1960s, he traveled to Washington for a variety of activities including functioning as a Roosevelt brain-truster, an advisor to Truman, a member of the National Security Council, a confidant of,

and later a speech writer for, Richard Nixon, and a state department advisor for both John F. Kennedy and Lyndon Johnson. His public and university involvement in Cold War politics and his position in international affairs have, in recent years, made him an object of criticism from several quarters as well as the *bête noire* of Lyndon Larouche.

*The Pragmatic Revolt in Politics: Syndicalism, Fascism, and the Constitutional State* (1928b) was an idiosyncratic book. Completed when Elliott was 31, it consisted in part of a prize essay at Balliol which was an element of his thesis under Lindsay (to whom the book was dedicated and who was an influential academic and public intellectual), but it incorporated, in somewhat repetitive form, several earlier articles (1922, 1924a, 1924b, 1925, 1926, 1927a). Although sympathetic to individuals (such as Burgess, Norman Wilde, Hocking, and R. M. MacIver), who in various ways sustained elements of the theory of the state, the work was neither a defense of the Germanic theory that had so long dominated the discourse of the discipline and the juristic conception of sovereignty associated with Austin nor a reflection of a one-dimensional ideology, but it did focus on the concept of the state as something more than government and on the conjunction of the ideas of democracy and community. Although distinctly problematizing pluralism, and less sympathetic to pluralism as a normative claim than were either Lindsay or Barker, upon whose work Elliott in part distinctly drew, his strategy was most generally informed by an attempt to accommodate the social reality of pluralism to the ideal of the state conceived as a people constituting a purposive organic entity. In this sense, the work resembled the argument of his Progressive contemporary, Mary Parker Follett (1918). In Kuhn's (1970) terms, it might be construed as an *ad hoc* exercise in shoring up a theory that was fast losing support. Like many of his English mentors, Elliott held on to remnants of German idealism while jettisoning much of the Hegelian and post-Hegelian baggage that Americans, particularly after the war, had begun to find distasteful and cumbersome. But most of Elliott's animus was directed toward the American philosophy of pragmatism which he confronted both as a particular argument, in the work of individuals such as William James and Dewey, and as a generic embrace of consequences over principles that he ascribed, in theory and practice, to everyone from Merriam to Mussolini. Pragmatism was, he claimed, the "*Zeitgeist*" of the era. In this respect, Elliott clearly recognized his enemy. It was in large measure the philosophy of pragmatism that spelled the end of the dominance of idealism and provided inspiration and support for both scientism and pluralism.

For Elliott, the "revolt" that he described and criticized had two symbiotically related dimensions: intellectual, and manifest in areas such as philosophy and social science; and political, characterized by movements such as syndicalism and fascism. Although he divided the book between an account of "pragmatic theory" and a description of "pragmatic politics," the themes were consistently blended. He counted

Bolshevism as part of the modern menace, but, he suggested, it was at least redeemed somewhat by its underlying rationalism which rendered it not “so skeptical of absolutes as pragmatism.” What Elliott was in part responding to was the aftermath of the post-Civil War retreat from abstract principles which was embraced by the founders of pragmatism (Menand 2001). What is ironic, or maybe instructive, is that it was in Harvard Circle that the philosophy of pragmatism most essentially took shape. Although rationalism in politics and in the study of politics had, Elliott claimed, seemed “ascendant” after the war, with the policies of Wilson and the advent of the League of Nations, there was, a decade later, in both theory and practice, and from both the ideological left and right, a “revolt against political rationalism” which entailed a rejection of the “rule of law” and the “constitutional and democratic state.” Although he recognized the political strains created by Reconstruction, “capitalistic industrialism,” and the “Great War” as contributing factors, the underlying problems, he maintained, were the growing influence of “voluntary associations,” or groups, and pragmatism which had provided both many contemporary political movements and social science with “their ideology and their values” (Elliott 1928b, vii).

Among those who influenced him, Elliott counted T. H. Green, MacIver, Leonard Hobhouse, R. B. Perry, Arthur Holcombe, A. N. Whitehead, and Hocking, but it was to Lindsay and, the Russian historian of English feudalism who had come to Oxford, Paul Vinogradoff, that he attributed his greatest debt. Lindsay had joined in criticizing the traditional doctrines of the state, but, while accepting much of English pluralism and socialism, he had held on to the image of the state as an inclusive purposive democratic unit and advocated a theory of constitutional sovereignty which, along with Barker’s ideas, was clearly manifest in Elliott’s reformulation. In a more negative vein, he specified Laski as his “greatest stimulant” (1928b, x), and this points to the extent to which the most crucial issue was not science but the theory and practice of democratic politics. Laski was a central figure in the introduction of pluralism into the conversation of American political science, and, for several years, he had been a resident scholar at the Harvard Law School and closely involved with pragmatists both in politics and in philosophy. Shortly before Elliott’s arrival, Lowell had sent Laski packing after his support of the Boston police strike, an event which loomed large in Elliott’s worries about the disestablishment of the constitutional state and the rise of pluralism.

In Elliott’s view, which was quite accurate, the state, as both concept and institution, was under attack, and he took this as tantamount to an attack on democracy. The state, he argued, was besieged from within by unions and other organizations, and externally the very idea of the state was rejected on the left by Communism and on the right by “Capitalistic Fascism” as well as by those who advocated a world-state. Although pluralist syndicalists attempted to “discredit” the authority of the state, fascist syndicalists sought to create an excessively unitary form, but both rejected the rule of

law and other attributes of liberal rationalism. Contemporary political science, Elliott claimed, had failed to discern these trends, because it had adopted the same basic pragmatic values and adapted and confined itself to a “scientific” description of the “facts” which was separate from values. Although Elliott believed that the field had not gone so far as to deny that a human being is a “purposive animal” with a capacity for reason, it had, he maintained, turned away from rationalism as a philosophy.

Elliott claimed that pragmatism had become so pervasive that it constituted the philosophy of both “revolt and reaction.” In political science, however, the effect had been simply to make the discipline “irrelevant.” Most fundamentally, he attributed this spirit of the age to the ideas of James and Dewey which had spilled over into political science and rendered the mainstream of the discipline “behavioristic in terms of psychology and positivistic in terms of philosophy.” This condition had both brought about a decline in the number of what he considered to be “true political theorists” and, as a consequence, a divide between political science and politicians who both, in their own way, had become infected with “pragmatic skepticism.” Although, he suggested, “absolute idealism” had been a mistake, the pragmatic reaction had gone too far as it infiltrated academic life and displaced the vision of such prophetic individuals as Walt Whitman. Even though he viewed pragmatism as a tributary of a larger anti-intellectualist trend in modern thought, it was most directly implicated in the “revolt against the sovereignty of the personalized state and against parliamentarianism” and stood behind not only syndicalism but also the “more chastened pluralism” of Laski, the “*droit objectif*” of Duguit, the “Fascist ‘efficiency’ gospel of Mussolini,” the doctrines of Sorel, and nearly every form of political and intellectual extremism (1928b, 7, 9).

Crucial to Elliott’s diagnosis was a provenance consisting of an extended dramatic story of the modern decline from the “age of reason” to the “age of skepticism” with latter’s rejection of rationalism and “universals.” Because “thought and act form a unity in history, this involved, in political terms, a focus on groups at the expense of both individuals and the state,” but group theory, he argued, “forms the rock upon which Idealism and pragmatism have alike gone aground with their ships of state” (1928b, 31–32). Although the former sought an extreme unity and ended in dictatorship, the latter, as in the case of the “general strike,” set class above state. The revolt against “intellectualism” and the triumph of “modernism,” represented by Dewey’s instrumentalism, was, Elliott argued, manifest in all aspects of culture and education from art to science. In political thought, this turn away from rationalism entailed a rejection of the theory of sovereignty that had, in its various forms, been put forth by Dicey, Burgess, W. W. Willoughby, Jellinek, and others. The consequence was that “the life of certain groups within the state, notably trade unions and professional associations, has become a more real thing in men’s experience than the common life represented by the state.” For Elliott, fascism was just the group idea taken to the

extreme—"if Hegel was the apologist of Prussianism, Duguit [with his reduction of law to power and social solidarity] is not less than of Fascism." Elliott claimed that while pragmatism may have had a leavening effect on the absolutist tendencies of idealism, it went too far and ended up providing a foundation for the "organically absolutist state of Fascist theory" (1928b, 40, 43, 64).

In political life, the pragmatic attitude and its abandonment of standards of value in favor of a consequentialist ethic had, Elliott argued, undermined stability and become the "mother of a brood of revolutionary theories of the state." He believed that he had, as an army officer, witnessed this tendency as he observed the May First demonstration in Paris during the armistice in 1919. Here, reason and law disappeared. The solution to this condition, he argued, was legal norms which provided an "accepted rule for fixing political responsibility." He argued that "constitutional government represents the same effort at political synthesis that conceptual logic does for thought synthesis. It must shun alike pluralism and absolutism" and stop the "centrifugal" tendencies of the former and curtail the "centripetal" forces of the latter. But, above all, constitutional government created a "community" as a "moral whole" in which popular sovereignty became a "reality" and to which the kind of "purpose" that supported democracy could be ascribed (1928b, 70, 75).

Another dimension of Elliott's work was a defense of what had become, in both the United States and England, the contested professional identity of political theory and an engagement of the concomitant issues of the identity of its object of analysis and the relationship between the political theorist and politics. Up to this point, few had ever explicitly suggested that there was any meaningful distinction between political science and political theory, and Merriam had sought to maintain their integration. As the theory of the state, and its philosophical grounds, waned and pluralism, with its emphasis on empirical methods, waxed, there was, as there would be a generation later, an emerging intellectual and professional tension between the general discipline and the subfield of political theory. Both sides of this divide were exercised, as they would continue to be, about the practical relationship between political science and politics, but although the division would persist for the remainder of the century and beyond, the practical concern, by the 1960s, would, in mainstream political science, fade to a discursive shadow. For Elliott, political theory was, rather than mere abstraction, a hybrid endeavor that required attention to facts, but it also consisted of a normative side with an emphasis on principles. Thus the "political theorist," he argued, was both a "political scientist" and a "political philosopher," and this combination was an intellectual mirror of the practical role of the "statesman" who must reconcile "means" and "ends." The problem, however, was that "there is not a single contemporary political theorist in America who is to be counted among those of the first order," and "most of our professors of politics would disdain the term *theorists*; they prefer to

call themselves political 'scientists'" (1928b, 5, 84, 85, 217).

Elliott was willing to admit that the concept of sovereignty, as conceived by individuals such as Austin, was a "relic," but that to view the state as merely one group among others, as Laski, Sorel, G. D. H. Cole, Duguit, and others had suggested, was to undermine the very idea of law. Sovereignty might in the end be a "fiction, but it was a necessary one for the survival of constitutional and democratic government. Elliott resurrected a version of the essential tenet of the nineteenth century theory of the state. This was the claim that the "state" referred to the People or sovereign community of which the government was only the agent. Elliott viewed the state as the organized form of the democratic community, but he maintained that the "government is the creature of the political community" and had only limited or delegated sovereignty as opposed to the complete sovereignty of the "federal state created by the Constitution." In his formulation, "the constitutional state . . . is the political community," which is in turn "a community of purpose" (1928b, 107, 247, 298). There can be no doubt that Elliott was by some measures what many would today think of as a Conservative, and he would increasingly move in that direction, but by other criteria, such as certain of his strictures against corporations, acts of government agents under color of law, and the autonomy of religious groups, he appeared quite differently. The core of his argument was that democracy required a national community, and, in this sense, he sounded much like Herbert Croly and other Progressives, and, despite his attack on Dewey, much like the book that Dewey (1927b) published in the previous year—*The Public and its Problems*.

Despite Elliott's rhetorical assimilation of the generic and specific images of pragmatism, as in his lumping together of James and Machiavelli, there were some actual, if tenuous, intellectual connections, but Elliott's claims about pragmatism as the root philosophy of Fascism and about Dewey's work as "an apology for the Fascist ideal of a 'disciplined' national organism" (1928b, 250) resembled the later equation of liberalism and totalitarianism in the work of individuals as diverse as Strauss and Max Horkheimer. On the whole, Elliott did quite a masterful job of weaving together his philosophical claims and his references to current political events, but like many of his generation, such as Friedrich, he was sadly mistaken about the Weimar constitution and in his prediction that, although many countries were abandoning representative democracy, "the new Germany seems steadfast in its practice of parliamentary government, under the benign moderation of Hindenburg" (1928b, 315). And his claims about syndicalist romanticism were in part drawn from the work of the German legal theorist Carl Schmitt who would inspire so many later, and now contemporary, political theorists on both the left and right—although he demurred with respect to Schmitt's critique of parliamentary government. Although Laski's pluralism was the focus of much of Elliott's critique, he, unlike Merriam was heartened by Laski's later efforts in the



*Grammar of Politics* (1925), which he concluded was a work of “political reconstruction” and “may well be the most important contribution that has been made to recent political theory” (1928b, 167). He may not have perceived exactly the direction (Marxist) in which Laski was moving, but, despite the residual pluralism, he was impressed by the new focus on state action and ethical issues.

Elliott supplemented his critique of pragmatism and pragmatist politics with his “co-organic” theory of groups. This was based on his essay written 6 years earlier in which he argued that a group possesses two kinds of consensus—one that is based primarily on economic activity and structured to attain its ends and another that consists of a moral agreement on values and purpose. It was, so to speak, both a *Gesellschaft* and *Gemeinschaft*. Thus, he claimed, groups are co-organic, much like an individual human personality. Political science, he claimed, was concerned primarily with the second form of consensus, that is, the “morally purposive element,” as manifest in the state. “A constitutional state is the product of a national community of political purpose as to the ends of the political association” (1928b, 355–56). A co-organic community, he argued, is not organic in the sense of constituting some superperson. Its will arises from the associated individual members, and although it moves beyond *laissez-faire* doctrines, it does not go to the extremes of communism and fascism. Economic pursuits can get out of hand, and, consequently, they often need to be regulated in order to bring them in line with political purpose. The co-organic state was not, he claimed, just a theoretical construction but had met the “pragmatic” test in the institutions of the United States and Britain, and although in certain respects it might, like sovereignty, be a myth, it was, he averred, a “true” or necessary myth.

Reponses to this young political theorist’s work were mixed. One reviewer noted, perceptively, that despite his opposition to critics of the state, “he is clearly on their side as against the old-fashioned simplification of political life” (C.D.B. 1929). Although some believed that the topic was important, the “clarity and coherence” of the argument were an issue (de Selincourt 1929). Another reviewer stressed the patchwork character of the book and stated that Elliott not only embraces a “vestigial idealistic structure in his thinking which operates to blind him to the facts” but also “fails utterly to understand Mr. Dewey’s philosophy and confuses pragmatism with a general concern with consequences which he himself shares” (Murphy 1929). Despite Elliott’s very positive review (1927b) of MacIver’s *The Modern State*, the latter was less generous. Although he noted that Elliott’s book was a worthy attempt to “redeem the long-continued neglect of political philosophy in the United States” and demonstrated how theories have “profound practical significance,” MacIver (1929) suggested that the diagnosis of pragmatism was “too generally applied” and that as far as reconciling pluralism and political order, the co-organic theory, despite its promise, was “not adequate to do so.”

## JOINING THE ARGUMENTS

The most lengthy and careful response to Elliott, however, was that of Catlin (1929) who characterized the work, somewhat ambiguously, as “indubitably the book of the year in political theory.” Catlin claimed that although “it is one of the most important expositions of what Professor Elliott himself calls the ‘new Liberalism,’ it ‘will be especially welcome in those constitutionally minded quarters which desire a new defense of the old Liberalism, with its belief in orthodox parliamentarianism or the permanence of the principles of the American Constitution, its stress on Constitutional action and its faith in the efficacy of reasonableness.’” Catlin acknowledged that his own “prejudice” was “entirely against the treatment of politics from the standpoint of value and purpose, if this treatment is to be considered not as supplementary but as the only satisfactory treatment.” Although Catlin agreed with the need for a “philosophy of political idealism,” he stressed that the practice of “political idealism” is the real problem” and that “realism is not opportunism.” It was necessary, he claimed, to recognize that “other factors enter into human life besides the conscious intellect so much belauded by Whigs and Liberals.” Because Elliott defended “the national state as the good, beautiful and the true,” he was actually the one who seemed “to be on the side of Mussolini” and failed to recognize that nationalism and internationalism are “deadly enemies” and that the “great issue of our time is whether we believe that sovereignty should ultimately reside in an international body representing that high good which is the ‘organized force of civilization’ or whether the national good is the *sum-mum bonum* and the national state the final sovereign” (Catlin 1929).

Elliott and Catlin again crossed paths in Stuart Rice’s influential *Methods in Social Science* (1931), which was sponsored by the Social Science Research Council and which was one of the most important documents of the Chicago School and the newly reconstructed American science of politics. Elliott was the dissenting voice in this collection where he sought once more to challenge the naturalistic image of political science and the pluralist account of politics advanced by individuals ranging from Hume to William B. Munro (1928) and Catlin. Although he did not object to the idea of a science of politics defined in some broad sense, he maintained that it was inappropriate to seek to apply the principles of natural science. Like Catlin, he preferred the term “Politics” as a designation of the field of study, because it not only semantically implied a bridge between theory and practice but also represented his image of “political theory” as encompassing “political philosophy and political science,” but he still saw the state as the principal object of analysis. His quarrel was in part with social science’s search for “universalized abstraction” and with Catlin’s attempt to create such an individualistic view of “political man” which, while supposedly represented in acts such as voting, was far from applicable to political systems in much of the contemporary world.



Although Elliott was not sympathetic to pluralism as a normative thesis, he did, like many later critics of pluralism, see group life as a fundamental reality, and problem, of modernity and as a "better field for a scientific attempt at examining the political act." Although it was possible to have a science of politics in the sense of a general commitment to objective description and comparison, he insisted on "resolutely renouncing any claim to discovering measurable variables in such a unit as political man (or woman)," because "the individual—given our present lack of a definitive scientific psychology—is still too unexplored and uncharted a realm to permit quantitative treatment of motives." It was, he claimed, in "groups of some permanence we get institutionalized behavior," but even here we are dealing with limited periods of time and particular cultural contexts. No one, for example, could, he argued, have predicted how Catholics would have reacted to the candidacy of Al Smith (who Elliott had supported). Moreover, what required study were the values and myths that moved people at particular times and places. "Timbucktoo cannot be governed like the island of Britain—though the government may offer some amusing parallels to that of Chicago, if one goes beyond forms to political realities" (Elliott 1931, 82, 86, 87, 91). In an "Appendix" to Elliott's article, Catlin in a modulated manner, continued to insist that the state was not the essence of politics and that political science should only be concerned with the particulars of "concrete reality" as instances of general laws grounded in human nature and that this included objects such as values about which generalizations could be adduced. He noted that "a study of values is indubitably quite as valuable as a study of social forces and controls," and although Elliott saw the typical generalizations of political science as inherently limited, Catlin preferred to see them as temporarily incomplete and not yet fully demonstrated (94).

In a later contribution, however, Catlin was more pointed in his criticism of Elliott and the dangers of a philosophical approach to the study of politics. He claimed that, although political philosophy and political science were legitimately both part of what Aristotle called *Politics*, political science should be "autonomous" and that to think that ethical philosophy can directly contribute to it rested on a "woeful confusion of means and ends or of values and existence." He suggested that because of intellectual lag most professional political scientists were still institutionalists who focused on the state and saw "all political theory" as a branch of ethics. This position, he argued, was being reinforced by individuals as diverse as Elliott and Laski as well as by some economists, and this explained why philosophy had been harmful to the study of politics. Even though ethics and the formulation of ends might be a part of the study of politics, the emphasis had too often been on positing ideals and absolutes which, when extended beyond the commitments of various groups to the society at large, bred intolerance. What Catlin recommended for political science was "practical agnosticism," but what is again crucial to note is that Catlin's emphasis on the separation of science and

ethics was for the purpose of gaining scientific authority that would be directed toward practical ends. It was a separation predicated on complementarity (Catlin 1933, 100, 114).

## CONCLUSION

Numerous things can be learned from a revisionist account of this period and from an understanding of the positions of Catlin and Elliott as, respectively, the leading and trailing edges of a crucial transformation. Although it is not possible here to elaborate on all the ways in which this conversation and the intellectual context in which it took place shaped, or might be construed as having shaped, the subsequent history of political science, it is important to note both that the similarities between the 1920s and behavioral era are not simply externally and retrospectively defined family resemblances and that the earlier period was not simply an adumbration. Part of what created forgetfulness regarding this period were the domestic and international crises which absorbed so much attention between 1929 and 1945, but in many respects, the behavioral "revolution" was the intellectual shadow of this earlier period. The exchange between Catlin and Elliott concretely initiated the form and content of debates about the character and role of political science that resurfaced during the behavioral movement as well as in more recent debates about democratic theory and about the character and role of political science.

Catlin's views prevailed in the mainstream discipline with respect to the nature of scientific theory and the search for a general theory of politics, the analytical demarcation and constitution of the domain of politics, the application of empirical and quantitative methods, the emphasis on interdisciplinary studies, the depreciation of historical and institutional research, the need for pure science to lead practical application, the separation of fact and value as well as the distinction between political science and political philosophy, and the wide acceptance of pluralism as an account of political reality and as a theory of democracy. He continued to make these arguments a generation later when they once again became an object of criticism and defense (e.g., 1956, 1957), and one would be hard pressed to find any basic tenet of behavioralism that was absent from Catlin's arguments. Whatever the diverse forms of research that individual political scientists would embrace after 1930, it would be more than two decades before there was any significant challenge to either the vision of science that Catlin so fully articulated or the theory of pluralist democracy that he defended. And when those challenges finally did arise in the subfield of political theory in the 1960s, they were markedly similar in form and content to arguments that had been mounted by Elliott who had continued to hold fast to his original position and apply it to a new context (e.g., 1940). One would be credulous to believe that these later arguments and the language in which they were couched were simply serendipitously discovered anew, that somehow the wide range of German émigré scholarship directed toward a criticism of American

political science just happened to look like the claims of Elliott or that the symmetry between Catlin's work and the behavioral program was only an interesting coincidence.

One thing that emerges clearly from an examination of the controversy in which Catlin and Elliott participated is the fact that the debate about the nature and value of scientific inquiry was secondary to theoretical and practical concerns. As in the case of the later debates about behavioralism, the prominence of the issue of science was misleading. Both Catlin and Elliott were dedicated professionally, and in their personal lives, as were leading figures in the next generation, to joining political science and practical politics, despite the fact that they embraced very different images of how to effect such a conjunction and of what it should achieve. Although Catlin believed that the authority of political science could be best secured by interdisciplinary borrowing and gaining legitimacy as science, Elliott believed that this was the threshold of a loss of disciplinary identity and political relevance. Lurking beneath the surface of the conversation were some distinct ideological and policy differences that rarely became explicit, but what most fundamentally separated them was a basic disagreement about democratic theory, and, as increased serious scholarly attention to the history of the discipline during the past two decades serves to remind us, the pursuit of science as well as the critique of that pursuit have never been disjoined from the search for the criteria and realization of democracy (Farr 2003; Smith 1997). What has been referred to as the "new" or "postbehavioral revolution" (Easton 1969), and more recently the sentiments represented by those involved in *Perestroika*, were largely reprises of these earlier concerns and commitments regarding the practical aims of the discipline. A closely related issue that was involved in the dispute, which would persist through the controversies of the 1960s, was the matter of professional differentiation within political science. It was in the course of the exchange between Catlin and Elliott that the first clear tension between political theory and empirical political science appeared, and the legacy of that exchange would persist and end in a fundamental professional and intellectual restructuring of the field.

Many of the main arguments in later controversies were in various ways directly connected to the work of Catlin and Elliott, but others, to this day, remain constrained by the discursive universe that these two individuals had played such an important role in establishing. It is always difficult to distinguish between, and assign relative weight to, personal and structural relationships. Some individuals, such as Strauss and Dahl, may have been less than fully aware of the intellectual and discursive heritage bequeathed by Elliott and Catlin, but in which they nevertheless significantly participated, while others such as Easton and Wolin, who maybe more than any other two individuals represented and articulated the poles of the controversy over behavioralism, were, in complicated ways, more directly connected. Although Wolin rejected Elliott's ideological position, his substantive vision of

democracy, his persistent critique of pluralism, his anti-positivism, his image of the declination of modern political thought and politics, and his defense of the autonomy of the "vocation" of political theory and its subject matter were all very similar to Elliott's arguments (1960; 1969). We might be reminded of Karl Mannheim's demonstration of how radicalism can arise from conservative romanticism, and, more to the point, how much Elliott's position in certain ways coincided with that of the Progressives. Although Easton never subscribed to the theory of pluralist democracy which was so widely associated with behavioralism, it was in part his negative predisposition against and reaction to Elliott's political views and depreciation of science that drew him theoretically and politically toward the work of the Chicago school and its image of the nature and goals of science, an image that had found its most articulate expression and defense in the work of Catlin. Despite Easton's quarrel with Catlin's idea of power as the organizing concept of political inquiry (which was common to the Chicago school in general), with his initiation of an early version of rational choice, and with his account of group equilibrium, Catlin's view of both the nature and purpose of scientific theory resonated in Easton's work (1953; 1969).

In the end, what an examination of the work of Elliott and Catlin most generally signifies is that as long as we continue to interpret the 1920s as an intimation of the 1950s, we will continue to read history backwards, and as long as we read history backwards, we will persist in misunderstanding important dimensions of our disciplinary identity and the genesis of issues and discursive forms that continue to inform the theory and practice of the field. This is not to say that an account of the history of the discipline should not be approached on the basis of purposes and perspectives formulated in the present but only that it is important to recognize, and seek edification from, the presence of the past.

## REFERENCES

- Almond, Gabriel. 1996. "Political Science: The History of the Discipline." In *A New Handbook of Political Science*, ed. Robert Goodin and Hans-Dieter Klingemann. Oxford: Oxford University Press, 50-96.
- Beard, Charles A. 1927a. "Time, Technology, and the Creative Spirit in Political Science" *American Political Science Review* 21 (February): 1-11.
- Beard, Charles A. 1927b. "Review of Catlin" *The Science and Method of Politics*. *American Political Science Review* 21 (August): 652-3.
- Catlin, G. E. G. 1925. "The Doctrine of Power and Party Conflict." *American Political Science Review* 19 (November): 718-34.
- Catlin, G. E. G. 1927a. "The Delimitation and Measurability of Political Phenomena," *American Political Science Review* 21 (May): 255-69.
- Catlin, G. E. G. 1927b. "Is Politics a Branch of Ethics?" *Monist* (July): 384-404.
- Catlin, G. E. G. 1927c. *The Science and Method of Politics*. New York: Knopf.
- Catlin, G. E. G. 1927d. "The Like-Minded Group." *The Philosophical Review* 36 (May): 237-52.
- Catlin, G. E. G. 1929. "Review of Elliott" *The Pragmatic Revolt in Politics*. *Political Science Quarterly* 44 (June): 259-65.
- Catlin, George E. G. 1930. *A Study of the Principles of Politics. Being An Essay Towards Political Rationalization*. London: George Allen and Unwin.

- Catlin, G. E. G. 1933. In symposium on "What Can Philosophy Contribute to the Study of Politics." *Aristotelian Society*, Supp. Vol. 12, 100–17.
- Catlin, G. E. G. 1935. "Review of H.D. Lasswell." *World Politics and Personal Insecurity Political Science Quarterly* 50 (June): 278–81.
- Catlin, George E. G. (1956) "The Future of Political Science." *Western Political Quarterly* 9 (December): 815–25.
- Catlin, George E. 1957. "Political Theory: What Is It?" *Political Science Quarterly* 72 (March): 1–29.
- Catlin, Sir George. 1972. *For God's Sake, Go!* Gerrards, Cross: Colin Smythe.
- C. B. D. 1929. Review of Elliott, *The Pragmatic Revolt in Politics*. *Journal of the Royal Institute of International Affairs* 8: 283.
- Crane, Robert T. 1927. "Review of Catlin." *The Science and Method of Politics. American Journal of Sociology* 33 (November): 483–5.
- Crick, Bernard. 1959. *The American Science of Politics*. Berkeley: University of California Press.
- Dahl, Robert. 1956. *A Preface to Democratic Theory*. Chicago: University of Chicago Press.
- Dewey, John. 1927a. "Politics and Human Beings." *New Republic* 50: 114–5.
- Dewey, John. 1927b. *The Public and Its Problems*. New York: Henry Holt.
- Dewey, A. Gordon. 1927c. "Review of Catlin." *The Science and Method of Politics. Political Science Quarterly* 42(): 617–21.
- Dickinson, John. 1930. "Democratic Realities and Democratic Dogma." *American Political Science Review* 24 (May): 283–309.
- Drysek, John S., and Stephen T. Leonard. 1988. "History and Discipline in Political Science." *American Political Science Review* 82 (December): 1245–60.
- Easton, David. 1953. *The Political System*. New York: Knopf.
- Easton, David. 1965. *A Framework for Political Analysis*. Englewood Cliffs: Prentice-Hall.
- Easton, David. 1969. "The New Revolution in Political Science." *American Political Science Review* 62 (December): 1051–61.
- Elliott, W. Y. 1922. "The Metaphysics of Duguit's Pragmatic Conception of Law." *Political Science Quarterly* 37 (December): 639–54.
- Elliott, W. Y. 1924a. "The Political Application of Romanticism." *Political Science Quarterly* 39 (June): 234–64.
- Elliott, W. Y. 1924b. "The Pragmatic Politics of Mr. H. J. Laski." *American Political Science Review* 18 (May): 251–75.
- Elliott, W. Y. 1925. "Sovereign State or Sovereign Group." *American Political Science Review* 19 (August): 475–99.
- Elliott, W. Y. 1926. "Mussolini, Prophet of the Pragmatic Era in Politics." *Political Science Quarterly* 41 (June): 161–92.
- Elliott, W. Y. 1927a. "An Ethics of Politics." *Political Science Quarterly* 42 (June): 259–67.
- Elliott, W. Y. 1927b. "Review of R. M. MacIver." *The Modern State. American Political Science Review* 21 (May): 432–4.
- Elliott, W. Y. 1928a. *Fugitives. An Anthology of Verse*. New York: Harcourt, Brace.
- Elliott, William Yandell. 1928b. *The Pragmatic Revolt in Politics: Syndicalism, Fascism, and the Constitutional State*. New York: Macmillan.
- Elliott, W. Y. 1931. "The Possibility of a Science of Politics: With Special Attention to the Methods Suggested by William B. Munro and George E. G. Catlin," with "Appendix: Commentary" by George E. G. Catlin, in *Methods in Social Science A Case Book*, ed. Stuart A. Rice, Chicago: University of Chicago Press, 70–94.
- Elliott, W. Y. 1940. "The Pragmatic Revolt in Politics: Twenty Years in Retrospect." *Review of Politics* 2.
- Ellis, Ellen Deborah. 1927. "Political Science at the Crossroads." *American Political Science Review* 21 (November): 73–91.
- Follett, Mary Parker. 1918. *The New State: Group Organizations, the Solution of Popular Government*. New York: Longmans, Green.
- Farr, James. 2003. "Political Science." In *The Cambridge History of Science*, vol. 7, *The Modern Social Sciences*, ed. Theodore M. Porter and Dorothy Ross. 306–28.
- Gosnell, Harold F. 1933. "Statisticians and Political Scientists." *American Political Science Review* 27 (June): 392–403.
- Gunnell, John G. 1993. *The Descent of Political Theory: The Genealogy of an American Vocation*. Chicago: University of Chicago Press.
- Gunnell, John G. 2004a. "The Real Revolution in Political Science." *PS. Political Science and Politics* 37 (January): 47–50.
- Gunnell, John G. 2004b. *Imagining the American Polity. Political Science and the Discourse of Democracy*. College Park: Pennsylvania University Press.
- Herring, Pendelton. 1940. *The Politics of Democracy*. New York: Norton.
- Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kuhn, Thomas. 1993. "Afterwords." In *World Changes: Thomas Kuhn and the Nature of Science*, ed. Paul Horwich. Cambridge: MIT Press, 311–39.
- Kuhn, Thomas. 2000. *The Road Since Structure*. Chicago: University of Chicago Press.
- Lasswell, Harold, and Abraham Kaplan. 1950. *Power and Society. A Framework for Political Inquiry*. Chicago: University of Chicago Press.
- MacIver, R. M. 1929. "Review of Elliott." *The Pragmatic Revolt in Politics. American Political Science Review* 23 (February): 194–6.
- Merriam, Charles E. 1931. Review of Catlin's *A Study of the Principles of Politics*. *International Journal of Ethics* 41 (April): 375–6.
- Menand, Louis. 2001. *The Metaphysical Club*. New York: Farrar, Strauss, and Giroux.
- Munro, William B. 1928. "Physics and Politics—An Old Analogy Revisited." *American Political Science Review* 22 (February): 1–11.
- Murphy, Arthur. 1929. Review of Elliott's *The Pragmatic Revolt in Politics*. *International Journal of Ethics* 39 (January): 239–44.
- Ricci, David. 1984. *The Tragedy of Political Science: Scholarship, Politics, and Democracy*. New Haven: Yale University Press.
- Ross, Dorothy. 1991. *The Origins of American Social Science*. Cambridge: Cambridge University Press.
- Sabine, George H. 1928. Review of Catlin, *The Science and Method of Politics. The Philosophical Review* 37 (May): 266–9.
- Seidelman, Raymond, with the assistance of Edward Harpham. 1985. *Disenchanted Realists: Political Science and the American Crisis, 1884–1984*. Albany: State University of New York Press.
- de Selincourt, O. 1929. Review of Elliott, *The Pragmatic Revolt in Politics. Mind* 38 (January): 106–10.
- Smith, Rogers M. 1997. "Still Blowing in the Wind: The American Quest for a Democratic, Scientific Political Science." *Daedalus* 126 (Winter): 253–87.
- Somut, Albert, and Joseph Tanenhaus. 1967. *The Development of American Political Science: From Burgess to Behaviorism*. Boston: Allyn & Bacon.
- Wolin, Sheldon S. 1960. *Politics and Vision: Continuity and Innovation in Western Political Thought*. Boston: Little-Brown.
- Wolin, Sheldon S. 1969. "Political Theory as a Vocation." *American Political Science Review* 62 (December): 1062–82.

# Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance

JANA VON STEIN | *University of Michigan*

**M**uch recent research has found that states generally comply with the treaties they sign. The implications of this finding, however, are unclear: do states comply because the legal commitment compels them to do so, or because of the conditions that led them to sign? Drawing from previous research in this Review on Article VIII of the IMF Treaty (Simmons 2000a), I examine the problem of selection bias in the study of treaty compliance. To understand how and whether international legal commitments affect state behavior, one must control for all sources of selection into the treaty—including those that are not directly observable. I develop a statistical method that controls for such sources of selection and find considerable evidence that the unobservable conditions that lead states to make the legal commitment to Article VIII have a notable impact on their propensity to engage in compliant behavior. The results suggest that the international legal commitment has little constraining power independent of the factors that lead states to sign.

**A**re international agreements only a reflection of states' preferences, or can they also alter leaders' interest in pursuing a particular course of action? In recent years, a number of international relations and legal scholars have sought to answer this question by examining whether states abide by the international legal commitments they make. Much of this literature has found that states generally comply with the treaties they sign, whereas enforcement problems are minimal (Chayes and Chayes 1995; Young 1994). As others have noted, however, compliance does not by itself demonstrate that international law constrains state behavior in meaningful ways. Downs, Rocke, and Barsboom (1996) argue that a state's decision to sign a treaty is endogenous to its expectations about future compliance. Consequently, compliance data alone do not tell us whether states abide by the treaties they sign because the legal commitment compels them to do so, or because they sign treaties that do not require significant departure from what they would have done in the absence of the treaty. To even begin to overcome this problem, one must first control for the basis of state selection (Downs, Rocke, and Barsboom, 383).

Theoretically and empirically, this insight is of central importance to the study of international institutions. Any theory of treaty compliance must recognize that institutional design is at least in part endogenous: states are only likely to invest their time and resources in agreements with which they have at least *some* interest in complying. This means that we must

also think about how the conditions that lead states to sign the agreement affect their postsigning behavior. Moreover, much of our reasoning must be expressed in counterfactuals: if the institution constrains state behavior, then it must be the case, all else equal, that a signatory would have engaged in compliant behavior less had it not signed, and/or that a nonsignatory would have engaged in compliant behavior more had it signed.

Empirical research on treaty compliance—both qualitative and quantitative—must also account for endogeneity and selection effects. This article explores the implications of these problems for the latter type of empirical research. If states sign international agreements only when certain conditions are present, examining whether signatories engage in compliant behavior more than do nonsignatories does not enable us to distinguish whether the behavior is attributable to the *agreement* itself, or to the *conditions* that led them to sign (Przeworski and Vreeland 2000, 387). One important way of mitigating this problem is by including in one's statistical analyses variables that control for the factors that affect both the decision to sign and the subsequent compliance. Yet, if some of these factors are *unobservable*, standard regression techniques will continue to yield biased results of the treaty commitment's effect.

Drawing from research in this Review on Article VIII of the International Monetary Fund (IMF) Treaty (Simmons 2000a), this article examines the problem of selection bias in the study of treaty compliance. I develop a statistical method that allows one to estimate the treaty commitment's effect on state behavior independent of *all* sources of selection—including those that cannot be directly measured. I find strong evidence that the unobservable factors that lead states to sign Article VIII significantly increase their propensity to engage in compliant behavior. The results with regard to nonsignatories are less conclusive, but suggest that the unobservable factors that lead states not to make the treaty commitment decrease their propensity to engage in compliant behavior. Failing to control for the sources of selection leads one to overstate considerably

Jana von Stein is Assistant Professor, Department of Political Science, University of Michigan, ISR, PO Box 1248, Ann Arbor, MI 48106 (janavs@umich.edu).

I am greatly indebted to Kenneth Schultz and Jeffrey Lewis for their invaluable assistance on every stage of this research. I also thank Beth Simmons for data and excellent comments; and Jeffrey Smith, Rob Salmond, Chad Rector, Barbara Koremenos, Dan Hopkins, Geoffrey Garrett, Lisa Blaydes, Neal Beck, Matthew Baum, and Nigel Ashford for very helpful comments. All remaining errors are my own. I gratefully acknowledge research support from the UCLA International Institute, the Institute on Global Conflict and Cooperation, the Institute for Humane Studies, and the Burkle Center for International Relations.

the effect of an Article VIII commitment on compliant behavior. Indeed, the international legal obligation appears to have little constraining power independent of the factors that lead states to sign.

### ARTICLE VIII COMMITMENT AND COMPLIANCE: PREVIOUS FINDINGS AND THE PROBLEM OF SELECTION BIAS

States that sign Article VIII of the IMF Treaty commit, among other things, to keeping the current account free from restriction. This entails allowing residents to use national currency or obtain foreign currencies to remunerate nonresidents for international transactions and permitting nonresidents who have obtained the national currency through current international transactions to use or transfer those balances (Edwards 1985, 390–93). Governments may wish to restrict the current account to mitigate balance-of-payments problems, or to support developmental goals that favor certain types of transactions (exports, capital inflows) over others (imports, capital outflows) (Simmons 2000a, 820). The Fund generally views these as undesirable practices that distort economies and hinder development (Edwards, 425–26).

Official IMF policy stipulates that while members may at any time inform the Fund that they accept the obligations of Article VIII, it is desirable that they “satisfy themselves that they are not likely to need recourse” to current account restrictions in the foreseeable future.<sup>1</sup> In practice, the Fund exercises significant discretion over the accession process. During annual consultations, it first encourages members that have not assumed Article VIII status to decrease or eliminate restrictions on the current account. Once a member has done so, the Fund usually then urges it to make the treaty commitment (Simmons 2000b, 581). In this manner, although the decision to sign ultimately lies in the hands of national authorities, the Fund’s Executive Board has been fairly successful at imposing its preference that a member not sign Article VIII until it has eliminated current account restrictions significantly or entirely (Edwards 1985 404, 422–23).

States cannot rescind an Article VIII commitment formally, and the IMF does not provide direct rewards for signing or punishments for not signing (Simmons 2000a, 823). Why, then, do states accept the treaty obligation? Simmons (819–21) argues, “Article VIII commitment is one way in which governments may seek to enhance their credibility to markets that doubt their ability or willingness to maintain current account policy liberalization . . . The acceptance of treaty obligations raises expectations about behavior that, once made, are reputationally costly for governments to violate.” In this interpretation, by signing Article VIII, governments attempt to signal their policy intentions by tying their hands—that is, by creating reputational costs that

they will suffer *ex post* if they renege.<sup>2</sup> This implies that signatories will be more likely to engage in compliant behavior, *ceteris paribus*. The analytical problem this poses, however, is that the *ceteris paribus* upon which the comparison hinges is unlikely to hold in practice: the IMF encourages countries it believes are ready to do so to sign, and the clearest indicator of such readiness is a low or null level of restrictions. If states sign only when certain conditions are present, it is difficult to distinguish whether signatories engage in compliant behavior more than do nonsignatories because of the *agreement* itself or because of the *conditions* that led them to sign (Przeworski and Vreeland 2000).

The intuition behind this problem can be clarified via a comparison from the field of medicine. Imagine that, to test the effectiveness of a new treatment, doctors ask sick people and healthy people to choose whether to take the drug, and then compare the health of those who took it with those who did not. In all likelihood, the sick will have opted to take the treatment in the hopes of being cured, whereas the healthy will have chosen not to do so because of potential side effects. If medical researchers attempt to draw conclusions about the drug’s effectiveness by comparing the two groups’ health, they will be unable to decipher whether the differences are attributable to the treatment or to the disease itself. Instead, of course, medical researchers test treatments by placing sick patients randomly into two groups—one that receives the treatment and another that is given a placebo. They can then draw unbiased conclusions about the drug’s effectiveness because they have two groups that are exactly alike, except that only one has received the treatment.

Just as it is not possible in the hypothetical medical example to determine the drug’s effectiveness by comparing the health of sick people who chose to take the treatment with that of healthy people who opted not to take it, it is not possible in the Article VIII case to draw conclusions about the treaty commitment’s constraining effect by comparing the restriction behavior of signatories to that of nonsignatories. Indeed, doing so does not tell us whether the observed behavior is attributable to the international legal commitment or to the underlying characteristics/conditions that lead states to sign or not sign. Yet in the Article VIII case, as in much social science research, we do not possess the experimental control that medical researchers do. We cannot create a control group of states that possess the attributes of nonsignatories but sign, or a control group of states that possess the attributes of signatories but do not sign. As a result, it is very unlikely that we will find two states that are alike in every way, except that one has signed and the other has not (Przeworski and Vreeland 2000, 386–87).

Hence, we are faced with a violation of one of the fundamental assumptions of classical regression theory: random selection. One important way of mitigating this problem is by controlling for the factors that

<sup>1</sup> Executive Board Decision 1034- (60/27), (*IMF Transitional Arrangements, Articles VIII and XIV*).

<sup>2</sup> See Fearon 1997 for game-theoretic models of signaling foreign policy interests using *ex post* or *ex ante* costs

affect both selection into the treaty (let us call this the selection equation) and the extent of compliant behavior (let us call this the outcome equation). Simmons makes important efforts to do so. Even when controlling for these sources of selection, she finds an Article VIII commitment to have a substantively large and statistically significant effect on restriction behavior. Indeed, signatories are up to 27% less likely to restrict the current account than are nonsignatories (Simmons 2000a, 830–31).

If, however, some *unobservable* factor(s) also leads states to sign and affects compliant behavior, estimates of the legal commitment's impact will continue to be biased.<sup>3</sup> In some instances, controlling for observed variables can *increase* the bias (Achen 1986; Przeworski and Limongi 1993). Indeed, although many of the conditions that lead states to sign agreements or undertake policies can be measured, some are unlikely to be measurable. Przeworski and Vreeland (2000, 387) and Vreeland (2002, 124) suggest, for example, that "political will" may affect a government's decision to enter an IMF program as well as its behavior subsequent to entering, but that this variable cannot be directly measured. Other examples of such unobservables include "trust" and "negotiation posture" (Vreeland 2003, 5–8, 52–54).

What unobservable factor(s) might affect commitment to and compliance with Article VIII? The IMF repeatedly has stated that by signing, a country "gives confidence to the international community that it will pursue sound economic policies." Similarly, Article VIII status is viewed by many as a "fundamental indicator of 'good standing' in the Fund."<sup>4</sup> A government's commitment to sound economic policies and/or desire to demonstrate "good standing" in the Fund are not directly observable attributes. Yet, these factors are likely to play a key role in determining a state's propensity to engage in compliant behavior and to accept Article VIII status. More specifically, because governments that place greater value on liberal economic policies and/or demonstrating "good standing" in the Fund are probably less likely to restrict, and those governments are probably also more likely to sign Article VIII, standard regression techniques are likely to overstate the extent to which being a signatory decreases the propensity to restrict. Conversely, because governments that place little value on liberal economic policies and "good standing" in the Fund are probably more likely to restrict, and those governments are probably also less likely to sign, standard regression techniques are likely to overstate the extent to which being a nonsignatory increases the propensity to restrict.

As I demonstrate formally later in this article, the result is that standard regression techniques are likely

to overstate the impact that a legal obligation to Article VIII has on restriction behavior, attributing to it the unobservable factors that lead states to sign or not sign and to engage in compliant behavior. Ideally, one would measure these unobservable attributes/conditions and include them in one's analyses. Because it is not likely that all sources of selection can be measured, we must instead adjust our statistical techniques (Przeworski and Vreeland 2000; Vreeland 2002, 2003).

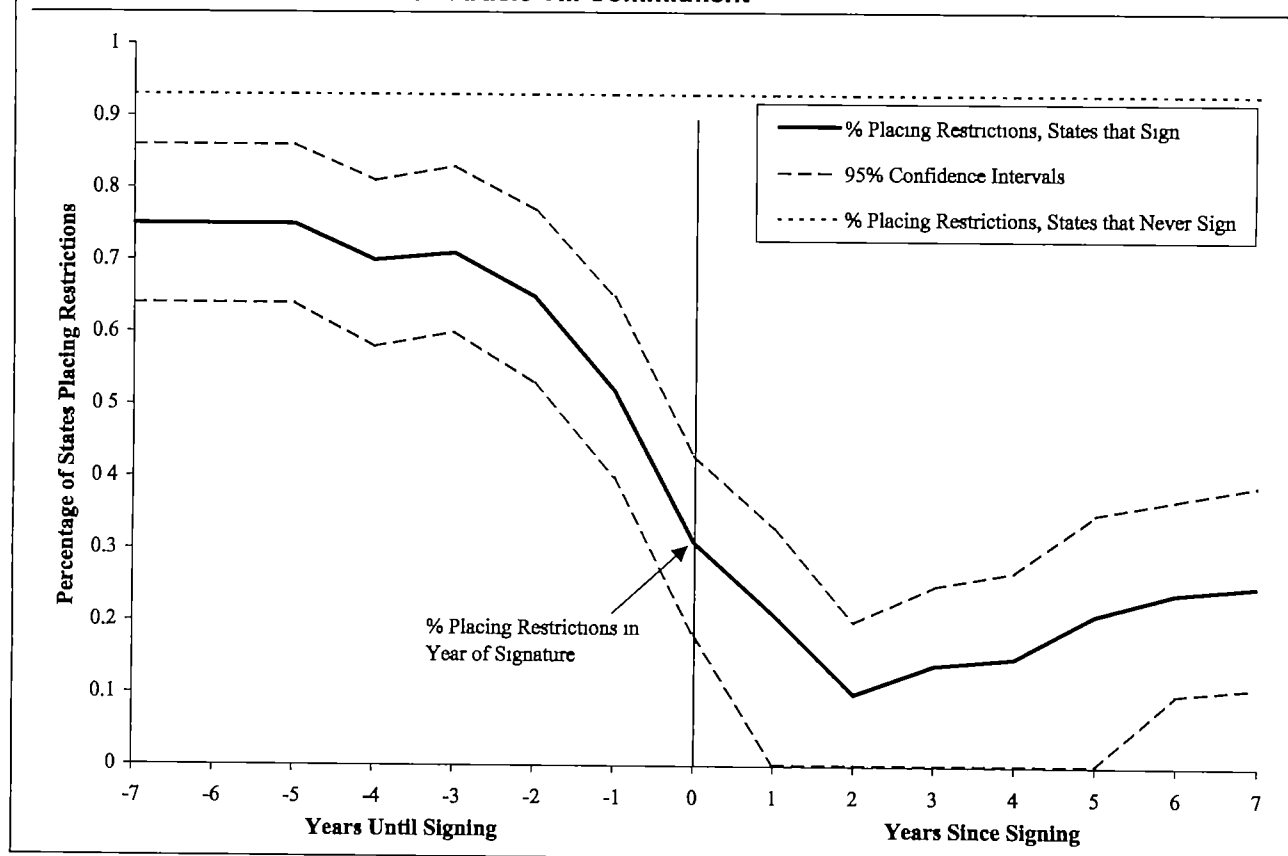
### WHO SIGNS? A CLOSER LOOK AT PATTERNS OF COMMITMENT AND COMPLIANCE

The previous section discusses why selection effects are likely to be present in the Article VIII case, and how one might expect them to affect estimates of the treaty commitment's impact on compliant behavior. I now turn to the empirical record and conduct a number of preliminary graphical and statistical analyses to determine whether there is evidence of selection and/or endogeneity. The data are yearly observations for up to 133 IMF members from 1967 to 1997 (Simmons 2000a). The dependent variable of interest, *Restrict*, equals one if a state placed restrictions in year  $t$ , and zero otherwise. The sample contains 1,354 signatory-observations (of which 350 restrict the current account) and 1,746 non-signatory-observations (of which 1,326 restrict the current account). Starting with those states that have not yet signed Article VIII but eventually sign, I calculate the average number of states placing restrictions as a function of the number of years remaining until signature. Next, for states that have already signed, I calculate the average number of states placing restrictions as a function of the number of years since signature. Finally, for states that never sign, I calculate the average number of states placing restrictions. Figure 1 displays these calculations, along with confidence intervals to assess the degree of variation in restriction behavior.

As Figure 1 demonstrates, a notable change in current account behavior takes place approximately 4 years prior to an Article VIII commitment: the percentage of countries placing restrictions decreases sharply from 70% to 31% and reaches levels considerably lower than those attained at any other point prior to signing. Immediately following the treaty commitment, the percentage of states placing restrictions continues to decrease; approximately 2 years after signing, however, the percentage of states placing restrictions *increases* somewhat. It is also of note that states that eventually sign or have already signed are *always* less likely to place restrictions than those that never sign ( $p < .05$ ). This suggests that there is something inherently different about Article VIII signatories, even decades before they sign. These preliminary observations neither confirm nor disconfirm the existence of unobservable sources of selection. They do, however, provide evidence that changes in restriction behavior precede signing, which suggests that selection into

<sup>3</sup> Simmons (2000a, 829) recognizes this possibility, but does not control for it.

<sup>4</sup> See IMF, "Zambia Accepts Article VIII Obligations," *Press Release No. 02/26*, 20 May 2002; and Shriraz Sidhva, "India Completes Key Reform of Currency," *Financial Times*, 21 August 1994, page 4, London Edition.

**FIGURE 1. The Percentage of States Placing Current Account Restrictions as a Function of the Number of Years to and since an Article VIII Commitment**

Article VIII is not random and that the treaty commitment is at least in part endogenous. A closer examination of restriction patterns is therefore in order.

To investigate with greater precision the patterns described previously, I examine three questions statistically. First, does the restriction behavior of states that are close to making an Article VIII commitment differ from their behavior long before signing? Clearly, one should expect to observe decreases in restriction levels as a state approaches the treaty commitment: the IMF encourages countries it believes are ready to do so to sign (Simmons 2000a, 820), and the clearest indicator of such readiness is a low or null level of restrictions. Second, does the restriction behavior of states that are close to signing Article VIII differ from the behavior of states that have already signed? If states are essentially behaving like signatories in the years leading up to an Article VIII commitment, this suggests that states may sign because they have reached low restriction levels, and not the opposite. Finally, do observable factors account fully for patterns of restriction behavior as a state approaches the treaty commitment?

To answer these questions, I conduct a probit analysis of the probability of current account restrictions.<sup>5</sup> The independent variables include those used by Simmons 2000a (including the variable *Article VIII*,

which equals one if a state has signed Article VIII, and zero otherwise) as well as two new independent variables. The first captures patterns of restriction behavior as a state approaches the treaty commitment. We must therefore determine at what point a nonsignatory should be considered “close to making an Article VIII commitment.” Figure 1 suggests that a substantial change in restriction behavior begins approximately 4 years prior to signing. Accordingly, I create the variable *Lead 4*, which equals one if a state will sign Article VIII in the next 1 to 4 years, and zero otherwise.<sup>6</sup> Because we are interested here in assessing restriction behavior as a state approaches an Article VIII signature, it is not entirely clear how one should code the year in which the state signs, henceforth referred to as  $t'$ . The most straightforward procedure is to create a second variable, *Year of Signature*, which equals one in year  $t'$  and equals zero for all other observations.<sup>7</sup>

first replicate Simmons's results, which are based on a logit model. I then utilize a probit model because the estimator employed later in this article relies for its starting values on the Heckman probit model. The logit and probit models yield similar results.

<sup>6</sup> Robustness checks suggest that the general result of a considerable decrease in the probability of restrictions in the years leading up to an Article VIII commitment holds across several codings of the *Lead* variable.

<sup>7</sup> This variable should be interpreted as the “added effect” of being in the year of signature because the Article VIII variable also equals one in year  $t'$ .

<sup>5</sup> See Simmons 2000a (833–34) for a description of the variables. For simplicity, I focus on the full models in Simmons 2000a (825, 830). I



I also include controls for temporal dependence.<sup>8</sup> Table 1 displays the results of the probit analysis.

The results reveal four interesting patterns. First, states that are within 4 years of signing are 18% less likely ( $p < .001$ ) to place restrictions than are other nonsignatories, *ceteris paribus*.<sup>9</sup> This suggests that states are not selected randomly into Article VIII—a finding that is sensible, given what we know about the Article VIII accession process, but that can have important consequences for the conclusions one draws about the effect of the treaty commitment on compliant behavior. Second, restriction behavior during the 4 years leading up to an Article VIII commitment is undistinguishable from that of states that have already signed (Wald test  $p$ -value = .865). This provides preliminary evidence that Article VIII status and compliance are at least in part endogenous. Third, states are considerably less likely to restrict in the year of signature than at any other time ( $p < .001$ ), all other observable variables being equal. This highly “virtuous” behavior during the year of signature appears to be a reflection of the fact that the Fund has generally not allowed members to move to Article VIII status and at the same time be placing restrictions (Edwards 1985, 422–23).

<sup>8</sup> Beck, Katz, and Tucker (1998) suggest that to control for temporal dependence in binary time-series-cross-section (BTSCS) data, the analyst use either a spell identification variable plus three splines, or a series of dummy variables marking the number of years since the last “event” (i.e., restriction). The data used in this article make the spline solution problematic for two reasons. First, the distribution of the spell identification variable is highly right-skewed. The STATA BTSCS routine (Tucker 1999) places the knots at the 25th, 50th, and 75th percentiles of the variable’s distribution: 0, 0, and 7 years since the last restriction. Two of the terms are therefore identical. A second problem is that the spell identification variable is distributed very differently for signatories than for nonsignatories, making it difficult to control for different patterns of temporal dependence in each group using the splines. This becomes particularly problematic when I implement the selection model later in this article, as separate outcome equations are estimated for signatories and nonsignatories.

The second solution proposed by Beck, Katz, and Tucker (1998) is preferable here because it is more flexible, and more adaptable to the skewed data distributions and different patterns of temporal dependence present in my data. A series of likelihood ratio tests suggests that dummies marking 0 and 1 years since the last restriction belong in the model, whereas dummies for subsequent years do not. To control for linear patterns of temporal dependence and to calculate the probabilities for Figure 2, I include the variable marking the number of years since the last restriction as well. Using temporal dummies rather than splines slightly decreases the Article VIII coefficient in the standard probit analysis, but not in a notable manner—the variable remains both statistically and substantively significant. In another analysis, I estimated the selection model outcome equations (see Table 2 and Figure 2) using splines rather than the temporal dummies. The results predict a slightly (6%) higher marginal effect for states that restricted in the previous year. Yet that analysis also predicts the international legal commitment to have an even smaller impact for states that are in their second year of current account liberalization than does the analysis using the temporal dummies. The results of the additional analysis are available from the author or at [www-personal.umich.edu/~janavapsr.html](http://www-personal.umich.edu/~janavapsr.html).

<sup>9</sup> I use *Clarify* (Tomz, Wittenberg, and King 2001) to estimate predicted probabilities and confidence intervals for the standard probit models in this paper. I use Gauss to estimate predicted probabilities and confidence intervals for the selection model. With both programs, I hold all other independent variables at their mean and vary the independent variable(s) of interest.

**TABLE 1. Results of Analysis of Current Account Restrictions as a State Approaches an Article VIII Commitment**

Independent Variables	Standard Probit Model 1
Lead 4 <sup>a</sup>	-.473*** (.116)
Year of Signature	-.931*** (.242)
Article VIII Signatory	-.494*** (.083)
Terms of Trade Volatility	.183*** (.054)
Balance of Payments/GDP	-.006* (.003)
Reserves/GDP	.357* (.179)
GDP Growth	-.012* (.006)
Use of IMF Credits	.364*** (.078)
Years since Last Restriction	-.034** (.012)
0 Years since Last Restriction	2.608*** (.128)
1 Year since Last Restriction	.384* (.180)
Constant	-1.726*** (.218)
Number of Observations	3,100
Log Likelihood	-693.440

*Note.* Figures are probit coefficients; robust standard errors are in parentheses. Dependent variable equals 1 if state restricted current account in year  $t$ , and 0 if not.

<sup>a</sup> Lead 4 equals 1 if state will sign Article VIII in next 1 to 4 years and 0 otherwise. \* $p < .05$ , \*\* $p < .01$ ; \*\*\* $p < .001$ .

Finally, the statistical significance of the Lead 4 variable suggests that some important patterns of variance in the dependent variable are not explained by the observable variables. Lead 4 may be proxying a number of unobservables that cause states to sign Article VIII and affect restriction behavior. Imagine, for example, that for one reason or another, a country’s leaders “convert” to the IMF’s economic orthodoxy. Their newfound commitment to liberal policies and to establishing/maintaining “good standing” in the Fund leads to a considerable change in current account policy, and, approximately 4 years later, to an Article VIII commitment. The Lead 4 variable may be proxying this unobservable variable(s).

An important caveat must be raised with regard to the findings discussed previously. The argument can be made that the observed decrease in restrictions as a state approaches an Article VIII commitment suggests not that states sign because they have reached low or null restriction levels, but instead that they cease to restrict the current account in the few years before signing because they are concerned about establishing in advance good postsigning reputations. The latter logic is plausible, and indeed the results displayed in Table 1 alone do not allow us to distinguish between the two interpretations. However, a number of additional

considerations suggest that the former and not the latter process is at work.

First, if the legal obligation did generate reputational costs that restricting nonsignatories would suffer after signing, one would expect the Fund to place greater emphasis on the importance of making a legal commitment to Article VIII, in an attempt to encourage nonsignatories to eliminate restrictions. Yet, the IMF's practices are much more suggestive of a selection process: the organization first focuses on promoting current account liberalization, and generally does not emphasize the international legal commitment until restrictions have been reduced significantly or eliminated. Second, if we believe that the observed shift in restriction behavior is attributable to states' desires to establish good postsigning reputations, we would not expect their restriction behavior to deteriorate after signing. Indeed, such behavior might expose them to criticism of being on good behavior solely in order to acquire Article VIII status. Yet as Table 1 indicates, states are considerably less likely to restrict in the year of signature than in subsequent years ( $p < .001$ ). It is difficult to believe that a state's good behavior before signing would carry it very far in the eyes of markets if it reimposed restrictions after committing to Article VIII.

Third, if postsigning reputational concerns affected restriction behavior, one would expect them to have an impact only during the period leading up to and following the treaty commitment. As a result, the restriction behavior of states that are far from signing but eventually sign should not differ systematically from that of states that never sign. Yet the empirical record reveals long-term differences in the two groups' restriction behavior, suggesting that there is something fundamentally different about signatories, even long before they sign.<sup>10</sup> This is more suggestive of a selection process whereby a certain "type" of state will assume the treaty obligation. Finally, if the reputational process were at play, one would expect Article VIII to constrain state behavior independent of the sources of selection. As we shall see in subsequent sections, however, there is little evidence that these constraining effects are present.

## A STATISTICAL MODEL OF TREATY COMMITMENT AND COMPLIANCE

The previous section provides evidence of the endogeneity of Article VIII. To understand why this is problematic for statistical inferences about the treaty commitment's effect on restriction behavior, it is necessary to examine why standard regression techniques

are likely to yield bias, and how the procedures originally proposed by Heckman (1976, 1979) and adapted to the particular observability problem described in the following present a solution. The Heckman probit model, which controls for sample selection when the outcome equation dependent variable is dichotomous (van de Ven and van Praag 1981), is common in political science research (e.g., Berinsky 1999; Lemke and Reed 2001). This model is appropriate for cases in which one observes the outcome of interest only for the selection group: for instance, one only observes whether states implement IMF austerity measures for the group of states that have committed to such measures (Vreeland 2002).

The Article VIII case presents a different partial observability problem: one only observes the restriction levels of Article VIII countries if they sign, and one only observes the restriction behavior of nonsignatories if they do not sign. If we believe that states that have signed differ in important ways from those that have not, then a signatory's decision on whether to restrict should be thought of as being fundamentally different from that of a nonsignatory. A decision to restrict is for a signatory a decision to *not comply* with a commitment, whereas for a nonsignatory, the issue of compliance is not part of a leader's calculus. It is therefore necessary to estimate three (rather than the standard two) equations: an equation determining selection into Article VIII, a noncompliance equation for signatories, and a restriction equation for nonsignatories. The techniques explained in the following and derived more fully in the Appendix do this.

Let the equation that determines selection into Article VIII be

$$\mathbf{z}^* = \mathbf{1}\theta + \mathbf{w}\gamma + \mu, \quad (1)$$

where  $\mathbf{z}^*$  is a state's latent propensity to sign;  $\mathbf{1}$  is an  $(n^S + n^N) \times 1$  vector of ones;  $n^S$  and  $n^N$  denote the number of observations for signatories and nonsignatories;  $\theta$  is the baseline propensity to sign;  $\mathbf{w}$  is an  $(n^S + n^N) \times k$  matrix of covariates that affect the probability of signing;  $\gamma$  is a  $k \times 1$  vector of coefficients; and  $\mu$  denotes the unobservable factors that determine a state's propensity to sign. I have conjectured that  $\mu$  includes unobservables that also make states less likely to restrict the current account, such as commitment to sound economic policies and/or desire to demonstrate "good standing" in the Fund. We do not observe  $\mathbf{z}^*$ . Rather, we observe  $\mathbf{z}$ , an  $(n^S + n^N) \times 1$  vector in which an element equals one if a state has signed Article VIII and zero if it has not.

Let the equation for whether a signatory places restrictions be

$$\mathbf{y}^S = \mathbf{1}\alpha^S + \mathbf{x}^S\beta^S + \varepsilon^S, \quad (2)$$

where  $\mathbf{1}$  is an  $n^S \times 1$  vector of ones;  $\alpha^S$  is a signatory's baseline propensity to restrict;  $\mathbf{x}^S$  is an  $n^S \times k$  matrix of covariates that affect the probability that a signatory will restrict;  $\beta^S$  is a  $k \times 1$  vector of coefficients; and

<sup>10</sup> In an additional analysis, I reestimated Model 1 and included a variable that equals one if a state is more than 4 years away from signing but eventually signs, and zero otherwise. The results confirm that states that are far from signing but eventually sign are significantly less likely ( $p < .001$ ) to restrict the current account than are states that never commit to Article VIII. The results of that analysis are available from the author or at [www.personal.umich.edu/~janavs/apsr.html](http://www.personal.umich.edu/~janavs/apsr.html).

$\varepsilon^S$  denotes the unobservable factors that determine a signatory's propensity to restrict, which I have hypothesized to be negatively correlated with  $\mu$  in equation (1). We do not observe  $\mathbf{y}^S$ . Instead, we observe  $\mathbf{y}^S$  (which equals zero if a signatory does not place restrictions and one if it does) *only* if  $\mathbf{z}$  equals one.

Let the equation for whether a nonsignatory places restrictions be

$$\mathbf{y}^N = \mathbf{1}\alpha^N + \mathbf{x}^N\beta^N + \varepsilon^N, \quad (3)$$

where  $\mathbf{1}$  is an  $n^N \times 1$  vector of ones;  $\alpha^N$  is a nonsignatory's baseline propensity to restrict;  $\mathbf{x}^N$  is an  $n^N \times k$  matrix of covariates that affect the probability that a nonsignatory will restrict;  $\beta^N$  is a  $k \times 1$  vector of coefficients; and  $\varepsilon^N$  denotes the unobservable factors that determine a nonsignatory's propensity to restrict, which I have hypothesized to be negatively correlated with  $\mu$  in equation (1). We do not observe  $\mathbf{y}^N$ . Rather, we observe  $\mathbf{y}^N$  (which equals zero if a nonsignatory does not place restrictions and one if it does) *only* if  $\mathbf{z}$  equals zero.

Now let us examine the standard probit model and why it may yield biased estimates of the effect of an Article VIII commitment on restriction behavior. A standard approach is to estimate the following:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}^S \\ \mathbf{y}^N \end{bmatrix} = \mathbf{1}\alpha^S + \mathbf{z}(\alpha^S - \alpha^N) + \begin{bmatrix} \mathbf{x}^S \\ \mathbf{x}^N \end{bmatrix} \beta + \begin{bmatrix} \varepsilon^S \\ \varepsilon^N \end{bmatrix}, \quad (4)$$

where  $\mathbf{1}$  is an  $(n^S + n^N) \times 1$  vector of ones;  $\alpha^N$  is a nonsignatory's baseline propensity to restrict;  $\mathbf{z}$  is an  $(n^S + n^N) \times 1$  vector whose elements equal one for signatories and zero for nonsignatories; and  $\alpha^S - \alpha^N$ , as previously defined, is the difference between a signatory's and a nonsignatory's baseline propensity to restrict.  $\mathbf{x}^S$  and  $\mathbf{x}^N$  are pooled together, forming an  $(n^S + n^N) \times k$  matrix of covariates that affect both signatories' and nonsignatories' propensity to restrict; and  $\beta$  is a  $k \times 1$  vector of coefficients.

As Heckman (1976, 1979) and others have shown, whether equation (4) will yield biased results, as well as the direction of the bias, in a function of the relationship between the unobservables that lead states to restrict and sign and the unobservables that lead states to restrict and not sign. In equation (4), any part of  $\varepsilon$  that is correlated with  $\mu$  will be attributed to  $\mathbf{z}$ . That is, standard techniques will attribute to being a signatory the unobservable shocks that affect both a state's propensity to restrict and its propensity to sign. Consider first what must be true if equation (4) is to yield unbiased results: the unobservables  $\varepsilon$  that affect a state's propensity to restrict are unrelated to the unobservables  $\mu$  that affect its propensity to sign:

$$\text{Cov}(\varepsilon^S, \mu) = 0; \text{ and } \text{Cov}(\varepsilon^N, \mu) = 0. \quad (5)$$

Here, equation (4) attributes no part of  $\varepsilon$  to  $\mathbf{z}$ . It produces consistent estimates of  $\alpha^S$  and  $\alpha^N$ , hence yielding unbiased estimates of the treaty commitment's effect.

Next, consider the case in which the unobservables  $\varepsilon$  that affect a state's propensity to restrict are positively correlated with the unobservables  $\mu$  that affect its propensity to sign:

$$\text{Cov}(\varepsilon^S, \mu) > 0; \text{ and } \text{Cov}(\varepsilon^N, \mu) > 0. \quad (6)$$

Standard techniques will attribute to  $\mathbf{z}$  any part of  $\varepsilon$  that is correlated with  $\mu$ . Because signatories on average have higher values of  $\mu$  than do nonsignatories, and because in equation (6)  $\varepsilon$  is positively correlated with  $\mu$ , standard techniques will overestimate  $\alpha^S$ , signatories' baseline propensity to restrict, and underestimate  $\alpha^N$ , nonsignatories' baseline propensity to restrict. As a result, standard techniques will in this case understate the treaty commitment's effect.

Finally, consider the case in which the unobservables  $\varepsilon$  that affect a state's propensity to restrict are negatively correlated with the unobservables  $\mu$  that affect its propensity to sign:

$$\text{Cov}(\varepsilon^S, \mu) < 0; \text{ and } \text{Cov}(\varepsilon^N, \mu) < 0. \quad (7)$$

Signatories on average have higher values of  $\mu$  than do nonsignatories, and in equation (7), these factors are negatively correlated with the unobservables  $\varepsilon$  that affect a state's propensity to restrict. As a result, standard techniques will underestimate  $\alpha^S$ , signatories' baseline probability to restrict, and overestimate  $\alpha^N$ , nonsignatories' baseline probability to restrict. Equation (4) will in this case overstate  $\alpha^S - \alpha^N$ , the treaty commitment's effect on restrictions. Because unobservables such as commitment to sound economic policies and/or desire to demonstrate "good standing" in the Fund that are thought to make states less likely to restrict are also thought to make them more likely to sign Article VIII, I hypothesize that the Article VIII case falls into this category.

To estimate the effect of Article VIII on restriction behavior independent of selection, I derived a likelihood function based on equations (1) through (3).<sup>11</sup>

The estimator developed here has two benefits in addition to controlling for the unobservable sources of selection. First, because it estimates the outcome equations for signatories and nonsignatories separately, it does not assume that the independent variables affect the restriction behavior of the two groups in the same manner (i.e., that  $\beta^S = \beta^N$ ). This alone does not necessitate controls for selection (a series of interaction terms would suffice), but given that I already intend to employ a selection model, the separate estimation of  $\beta^S$  and  $\beta^N$  provides additional flexibility. Second, the estimator developed here allows the outcome equations to contain different columns of  $\mathbf{x}^S$  and  $\mathbf{x}^N$ . If we believe that those states that have made a treaty commitment differ in important ways from those that have not, it is also likely that some independent variables affect one

<sup>11</sup> See Heckman 1976, 1979; and van de Ven and van Praag 1981 for further details. The Appendix contains the statistical proofs, the likelihood function, and an explanation of the identification restrictions. The STATA code is available from the author or at [www-personal.umich.edu/~janavv/apsr.html](http://www-personal.umich.edu/~janavv/apsr.html).

**TABLE 2. Results of Analyses of the Probability of Current Account Restrictions**

Independent Variables	Standard Probit Model 2	Selection Model, Signatories	Selection Model, Nonsignatories
Article VIII	-.532*** (.079)	—	—
Terms of Trade Volatility	.196*** (.054)	.262*** (.082)	.153* (.077)
Balance of Payments/GDP	-.006* (.003)	-.002 (.003)	-.011 (.006)
Reserves/GDP	.233 (.169)	-.526 (.609)	.628* (.281)
GDP Growth	-.013* (.006)	-.022* (.011)	-.011 (.008)
Use of IMF Credits	.364*** (.076)	.511*** (.119)	.277*** (.111)
Years since Last Restriction	-.035*** (.012)	-.037* (.016)	-.036 (.019)
0 Years since Last Restriction	2.542*** (.127)	2.267*** (.185)	2.807*** (.187)
1 Year since Last Restriction	.334 (.178)	.148 (.290)	.565* (.248)
Constant	-1.771*** (.216)	-2.229*** (.306)	-1.887*** (.332)
$\rho$	—	-.533*** (.153)	-.339 (.224)
Number of Observations	3,100	1,288	1,684
Log Likelihood	-709.130	-834.371	

*Note.* Figures are probit coefficients, robust standard errors are in parentheses. Dependent variable equals 1 if state restricted the current account in year  $t$  and 0 if not.  $\rho$  measures sample selection and can assume values from  $-1$  to  $+1$ .

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

group's decision to engage in compliant behavior and not the other's. This is therefore an additional source of flexibility in statistical estimation.

An important concern remains with regard to the estimator I use. The selection equation examines why states sign Article VIII; hence, the estimates should be based on the independent variables' effects before and when states sign, but not after. Because once an Article VIII commitment is made, it cannot be rescinded, survival analysis techniques that focus on the spell of time until signing occurs are necessary for the selection equation (Simmons 2000a, 823). Yet, the selection model requires a probit model for the selection equation. This problem can be circumvented by creating a dummy variable which equals one for all observations after year  $t'$ , and zero otherwise. When this dummy variable is included in the probit equation, the estimated coefficients, standard errors and z-scores of the independent variables are based only on the values of the independent variables before or in year  $t'$  (i.e., prior to signing and in the year of signing). This makes it possible to estimate a probit model in the selection equation while still accounting for the nature of the data.<sup>12</sup> To ensure that the transformation from a Cox

to a probit model is not accounting for the differences between my results and Simmons's, I confirm that the two models yield comparable estimates.<sup>13</sup>

## RESULTS

The results of the statistical analysis controlling for selection appear in Table 2. For comparative purposes, Table 2 also displays the standard probit model's estimates. The results provide strong evidence of selection effects. Indeed, a likelihood ratio test that the joint effect of the correlation coefficients  $\rho^S$  and  $\rho^N$  equals zero is highly significant ( $p < .001$ ), suggesting that the selection model employed here maximizes the likelihood function significantly better than do methods not controlling for selection.  $\rho^S$  is negative and highly statistically significant ( $p < .001$ ), indicating as hypothesized that the unmeasured conditions that lead states to commit to Article VIII make them

nonsignatories do not pose a problem here, and because the variable marking the number of years since joining is almost perfectly normally distributed.

<sup>13</sup> Using the coefficients generated by the probit analysis, I calculate for each independent variable the "relative risk" of signing; that is, the ratio of the predicted probability of signing when there is a one-unit increase in the independent variable to the predicted probability of signing before the one-unit increase. These "relative risks" can be directly compared with the Cox model coefficients. All other variables are held at their mean. The results are available from the author or at [www-personal.umich.edu/~janavs/aprs.html](http://www-personal.umich.edu/~janavs/aprs.html).

<sup>12</sup> To control for temporal dependence, I include a variable marking the number of years since the state joined the IMF (which is precisely the Cox function's "time until failure") as well as three cubic splines (Beck, Katz, and Tucker 1998). I employ this approach because differences in patterns of temporal dependence among signatories and

considerably less likely to restrict the current account. As conjectured,  $\rho^N$  is negative, providing evidence that the unobservable factors that cause states not to sign make them more likely to restrict the current account.  $\rho^N$  falls short of conventional levels of statistical significance ( $p = .130$ ), and therefore one cannot reject the null hypothesis that there are no selection effects for nonsignatories. Note, however, that standard regression techniques will yield biased results as long as selection effects are present for at least one of the two groups.<sup>14</sup>

That selection effects are stronger and more systematic for signatories may also be sensible given how the Article VIII accession process works. Because it is generally the Fund that urges members to sign (Simmons 2000b, 581), noncompliant "types" are unlikely to be willing or able to attain the low or null restriction levels necessary to be approached by the IMF and "encouraged" to sign. On the other hand, because the Fund cannot obligate members to commit to Article VIII (Gold 1988, 227), some compliant "types" will choose for one reason or another to delay accepting Article VIII status. Hence, it is likely that there are fewer noncompliant "types" that have signed than compliant "types" that have not signed. In other words, the IMF is probably more successful at screening out bad apples than it is at forcing good apples to sign. This apparent "asymmetrical selectivity" may explain why selection effects are stronger and more systematic for signatories than for nonsignatories.

Another important test of Article VIII's independent effect is whether—all else equal—the international legal commitment has a strong negative impact on the probability of restrictions (Simmons 2000a, 830). The techniques implemented here make possible the estimation of such counterfactuals by holding constant for all other conditions—including those that cannot be directly measured. This involves two steps. First, take the average nonsignatory and estimate its probability of restricting when it is exposed to the conditions—observable and unobservable—to which nonsignatories are exposed. To do so, I estimate the probability of restrictions as predicted by the selection equation and the nonsignatories' outcome equation, using the mean values of the nonsignatories' independent variables. Second, take the same average nonsignatory and determine what its probability of restrictions would have been had it signed. To do

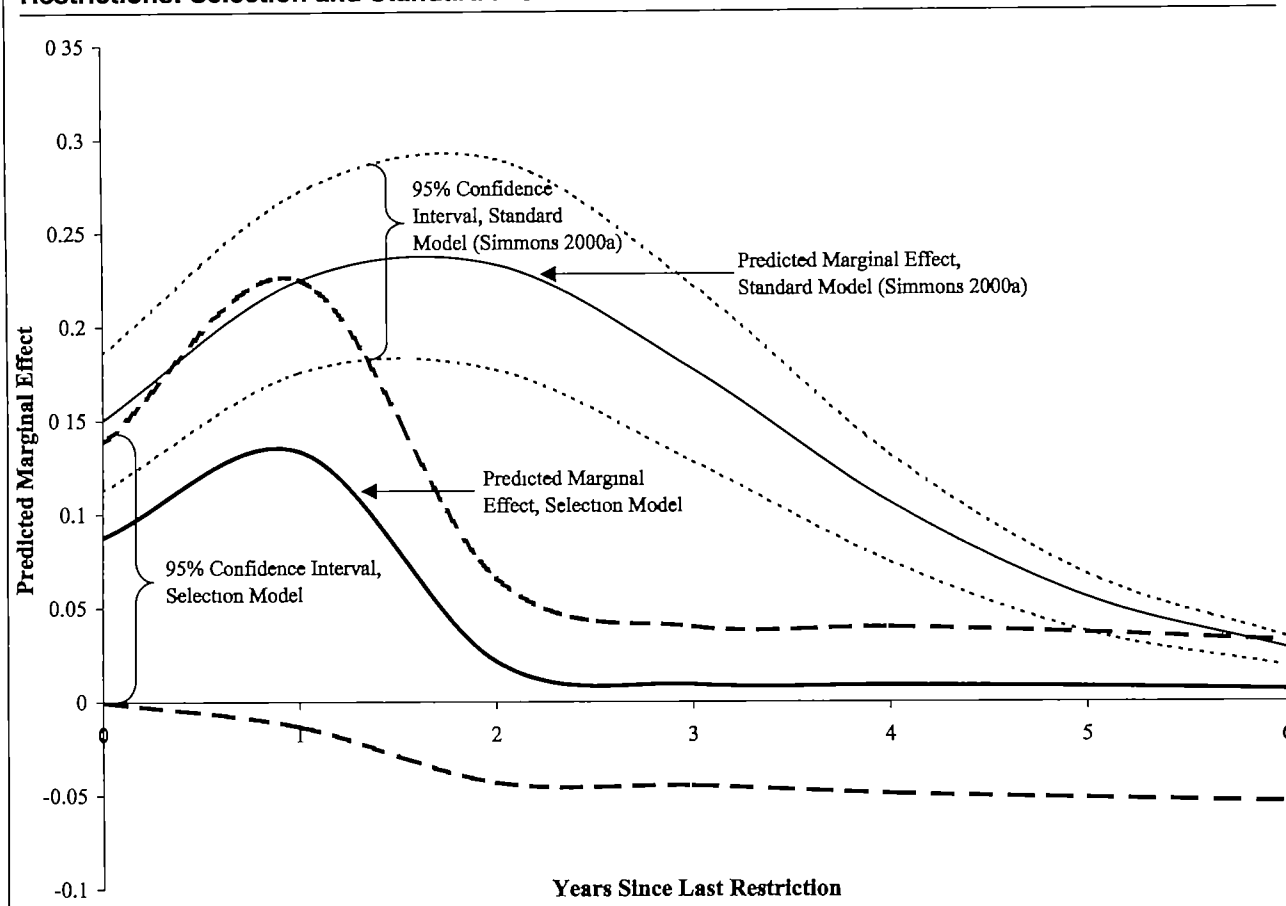
so, I use the same mean values of the nonsignatories' independent variables and estimate the probability of restrictions as predicted by the selection equation and the signatories' outcome equation. We now know the probability of current account restrictions for two countries that are alike in every way, except that the latter has made a legal commitment to Article VIII and the former has not. The difference between these probabilities yields the marginal effect of an Article VIII commitment, independent of selection. Figure 2 displays these marginal effects, as well as those produced by Simmons's (831) standard probit model, as a function of time elapsed since the last restriction.

The previous analysis yields two important results. First, failing to control for the unobservable sources of selection consistently overstates the effect of an Article VIII commitment on restriction behavior. Indeed, selection bias accounts for between 31% and 95% of the standard probit model's estimated effect of the legal commitment on a state's propensity to engage in compliant behavior. Second, once one controls for the unobservable sources of selection, the treaty obligation is found to have only a limited independent effect on state behavior. For states that are in their first year of current account liberalization, the legal commitment does constrain: signatories are 13% less likely than are nonsignatories to restrict the current account, *ceteris paribus* ( $p < .05$ ). Subsequently, however, the treaty commitment's effect virtually disappears. By the second year of liberalization—when the standard probit model estimates an Article VIII commitment to matter the most, making signatories 23% less likely to restrict the current account than nonsignatories (Simmons 2000a, 831)—methods controlling for selection suggest that the legal commitment has virtually no independent effect on state behavior.

The results also confirm that estimating the impact of the observable variables separately for signatories and for nonsignatories significantly maximizes the likelihood function ( $p < .05$ ). A series of Wald tests indicates the following. Volatility in the terms of trade and GDP growth affect restriction behavior in approximately the same manner for signatories and nonsignatories. Increases in the balance of payments as a proportion of GDP have a stronger negative effect on the probability of restrictions for nonsignatories than for signatories. Article VIII signatories that use Fund credits are more likely to restrict than are nonsignatories that use credits, *ceteris paribus*. Increases in reserves as a proportion of GDP appear to decrease the probability of restrictions for signatories (though not at standard levels of statistical significance), whereas they increase the probability of restrictions for nonsignatories ( $p < .05$ ).<sup>15</sup> The most considerable difference between the two groups concerns patterns of temporal dependence. All else equal, nonsignatories that are within

<sup>14</sup> Consider the case in which negative selection effects exist for the signatory group but not for the nonsignatory group:  $\rho^S < 0$  and  $\rho^N = 0$ . Suppose that the biased (standard probit)  $\alpha^S$  coefficient = 1, while the unbiased  $\alpha^S$  coefficient = 2 (the numerical values are hypothetical, but their ordering makes sense substantively). Because we are assuming here that there are no selection effects for nonsignatories, suppose that in both the standard and selection models,  $\alpha^N = 2$ . The standard probit model would estimate  $\alpha^S - \alpha^N = -1$ . Controlling for selection, however,  $\alpha^S - \alpha^N = 0$ . Therefore, standard techniques would yield biased estimates of  $\alpha^S - \alpha^N$  even if  $\rho^N = 0$ . Clearly, the stronger the selection effects for both groups (i.e., as  $\rho^S$  and  $\rho^N \rightarrow -\infty$ ), the more standard techniques will overstate (in the negative direction)  $\alpha^S - \alpha^N$ . Yet, relatively large negative values of  $\rho^S$  will also lead one to overstate the extent to which signing decreases a state's propensity to restrict even if  $\rho^N = 0$ .

<sup>15</sup> This result is somewhat perplexing. However, it is consistent with Simmons's (2000a) findings. Reestimation of the model without this variable in the outcome equations does not change the results notably.

**FIGURE 2. The Marginal Effect of an Article VIII Commitment on the Probability of Current Account Restrictions: Selection and Standard Probit Models**

one year of their last restriction are notably more likely to place restrictions ( $p < .05$ ) than are signatories that are within one year of their last restriction. After the first year, however, the difference is no longer significant.

## CONCLUSION

This article has demonstrated that selection effects can have important consequences for the conclusions we draw about the impact of treaty commitments on state behavior. I have shown that the unobservable conditions that lead states to sign Article VIII of the IMF Treaty make them considerably more likely to engage in compliant behavior. I have also found evidence, albeit less conclusive, that the unmeasured factors that cause states not to commit to Article VIII make them less likely to engage in compliant behavior. Failing to control for selection effects leads one to overstate considerably the extent to which the treaty obligation affects states' restriction behavior. Indeed, if the conditions that led a state to sign change, a legal commitment to Article VIII appears to have little constraining power.

Although this study has examined one article of one treaty, it presents a methodological challenge to schol-

ars conducting empirical research on treaty compliance more broadly. It demonstrates that a fundamental part of understanding whether and why states comply lies in understanding what drives behavioral change as a state approaches a treaty commitment, and how changes in those conditions affect subsequent compliance. For scholars employing quantitative methods, the central implication of this article is that in order to obtain unbiased estimates of the treaty commitment's impact on state behavior, statistical methods that control for the unobservable sources of selection are very likely to be necessary. For scholars using qualitative methods, the chief implication is that it is important to consider not only the extent of compliant behavior both after and well before signature but also what drives the decision to sign (or not sign) and determines the extent of compliant behavior.

My findings also have some interesting substantive implications for how we think about treaty commitment and compliance. The results cast doubt on the argument that an Article VIII obligation serves as a constraining mechanism that raises the reputational costs a state will pay if it reneges (Simmons 2000a, 819). Why, then, do states sign? That is, if the decision to commit is largely endogenous to expectations about future compliance (Downs, Rocke, and Barsoom 1996), why do states sign at all? Article VIII may instead serve

as a screening device. In this conception, Article VIII status does, as Simmons (819–21) argues, signal future policy intentions to markets, which possess incomplete information. However, the mechanism at work is different from that suggested by Simmons. Signing Article VIII enables leaders credibly to signal their intention to engage in compliant behavior in the future not because the legal commitment generates *ex post* reputational costs for noncompliance, but because the *ex ante* costs of becoming a signatory are high enough to deter non-compliant “types” from signing. If the political capital and effort (formal or informal) necessary to become a signatory are sufficiently costly *ex ante*, states are likely to comply because the requirements for entry effectively screen compliant “types.”

This article’s findings may lead some to adopt the bleak view that—at least with regard to Article VIII of the IMF Treaty—international institutions do little or nothing to promote compliant behavior. I believe the evidence points toward a different interpretation. In the Article VIII case, a central role of the Fund appears to lie not in advocating the legal commitment itself, but in promoting—both before and after signature—the conditions that lead states to make treaty commitments and to engage in compliant behavior. Another fundamental role lies not in monitoring and punishing defectors, but in using formal and/or informal requirements for entry to screen potential signatories. Different international cooperation problems call for different institutional solutions (Koremenos, Lipson, and Snidal 2001), and it is not the claim of this article that all international institutions fulfill functions similar to those I have identified in the Article VIII case. Under what conditions do international institutions play these roles rather than others? In what circumstances is the prospect of signatory status sufficient to compel states to become compliant “types?” These questions point to interesting new avenues of thinking about the nature of interactions between states and international institutions.

## APPENDIX: DERIVATION OF THE LIKELIHOOD FUNCTION

Let  $\rho^S = \text{Cov}(\varepsilon^S, \mu)$ ;  $\rho^N = \text{Cov}(\varepsilon^N, \mu)$ . Let  $\Phi_2$  denote the standard bivariate normal cumulative distribution function.

For signatories, the probability of not restricting is

$$\begin{aligned} \Pr(y^S = 0) &= \Pr(z = 1, y^S = 0) \\ &= \Pr(w\gamma + \mu > 0 \text{ and } x^S\beta^S + \varepsilon^S < 0) \\ &= \Pr(\mu > -w\gamma \text{ and } \varepsilon^S < -x^S\beta^S) \\ &= \Phi_2(w\gamma, -x^S\beta^S, -\rho^S) \end{aligned} \quad (8)$$

For signatories, the probability of restricting is

$$\begin{aligned} \Pr(y^S = 1) &= \Pr(z = 1, y^S = 1) \\ &= \Pr(w\gamma + \mu > 0 \text{ and } x^S\beta^S + \varepsilon^S > 0) \\ &= \Pr(\mu > -w\gamma \text{ and } \varepsilon^S > -x^S\beta^S) \\ &= \Phi_2(w\gamma, x^S\beta^S, \rho^S) \end{aligned} \quad (9)$$

For nonsignatories, it is straightforward that the probability of not restricting is

$$\Pr(y^N = 0) = \Phi_2(-w\gamma, -x^N\beta^N, \rho^N) \quad (10)$$

and that for nonsignatories, the probability of restricting is

$$\Pr(y^N = 1) = \Phi_2(-w\gamma, x^N\beta^N, -\rho^N) \quad (11)$$

The likelihood function is as follows:

$$\begin{aligned} L &= \left[ \prod_{y^S=1} \Phi_2(w\gamma, x^S\beta^S, \rho^S) \right] \left[ \prod_{y^S=0} \Phi_2(w\gamma, -x^S\beta^S, -\rho^S) \right] \\ &\times \left[ \prod_{y^N=1} \Phi_2(-w\gamma, x^N\beta^N, -\rho^N) \right] \\ &\times \left[ \prod_{y^N=0} \Phi_2(-w\gamma, -x^N\beta^N, \rho^N) \right] \end{aligned} \quad (12)$$

In order for equation (12) to be identified,  $x$  must contain at least one variable not contained in  $w$ , or  $w$  must contain at least one variable not contained in  $x$ . The derivations of equations (10) and (11) and the STATA code for equation (12) are available from the author or at [www-personal.umich.edu/~janavs/aprsr.html](http://www-personal.umich.edu/~janavs/aprsr.html).

## REFERENCES

- Achen, Christopher. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.
- Beck, Nathaniel, Jonathan Katz, and Richard Tucker. 1998. “Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable.” *American Journal of Political Science*, 42 (October): 1260–88.
- Bernsky, Adam. 1999. “The Two Faces of Public Opinion.” *American Journal of Political Science* 43 (October): 1209–30.
- Chayes, Abram, and Antonia Handler Chayes. 1995. *The New Sovereignty: Compliance with International Regulatory Agreements*. Cambridge: Harvard University Press.
- Downs, George, David Rocke, and Peter Barsoom. 1996. “Is the Good News about Compliance Good News about Cooperation?” *International Organization* 50 (Summer): 379–406.
- Edwards, Richard. 1985. *International Monetary Collaboration*. Dobbs Ferry, NY: Transnational Publishers.
- Fearon, James. 1997. “Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs.” *The Journal of Conflict Resolution*, 41 (February): 68–90.
- Gold, Joseph. 1988. *Exchange Rates in International Law and Organization*. Chicago: American Bar Association.
- Heckman, James. 1976. “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models.” *Annals of Economic and Social Measurement* 5 (4): 475–92.
- Heckman, James. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47 (January): 153–62.
- International Monetary Fund. Archives S424. *Transitional Arrangements, Articles VIII and XIV*. Various years.
- Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. “The Rational Design of International Institutions.” *International Organization* 55 (Fall): 761–99.
- Lemke, Douglas, and William Reed. 2001. “War and Rivalry among Great Powers.” *American Journal of Political Science* 45 (April): 457–69.
- Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- Przeworski, Adam, and Fernando Limongi. 1993. “Political Regimes and Economic Growth.” *Journal of Economic Perspectives* 7 (Summer): 51–69.



- Przeworski, Adam, and James Vreeland. 2000. "The Effect of IMF Programs on Economic Growth" *Journal of Development Economics* 62 (August): 385-421.
- Simmons, Beth. 2000a. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs" *American Political Science Review* 4 (December): 819-35.
- Simmons, Beth. 2000b. "The Legalization of International Monetary Affairs." *International Organization* 54 (Summer): 573-602.
- Tomz, Michael, Jason Wittenberg, and Gary King. 2001. "Clarify: Software for Interpreting and Presenting Statistical Results." <http://gking.harvard.edu/clarify> (June 27, 2003).
- Tucker, Richard. 1999. "BTSCS: A Binary Time-Series Cross-Section Data Analysis Utility." <http://www.vanderbilt.edu/rtucker/programs/btscs/>. (June 27, 2003).
- van de Ven, Wynand, and Bernard van Praag. 1981. "The Demand for Deductibles in Private Health Insurance: a Probit Model with Sample Selection." *Journal of Econometrics* 17 (November): 229-52.
- Vreeland, James. 2002. "The Effect of IMF Programs on Labor." *World Development* 30 (January): 121-39.
- Vreeland, James. 2003. *The IMF and Economic Development*. Cambridge: Cambridge University Press.
- Young, Oran. 1994. *International Governance: Protecting the Environment in a Stateless Society*. Ithaca, NY: Cornell University Press.

# The Constraining Power of International Treaties: Theory and Methods

BETH A. SIMMONS *Harvard University*  
DANIEL J. HOPKINS *Harvard University*

**W**e acknowledge the contribution of von Stein (2005) in calling attention to the very real problem of selection bias in estimating treaty effects. Nonetheless, we dispute both von Stein's theoretical and empirical conclusions. Theoretically, we contend that treaties can both screen and constrain simultaneously, meaning that findings of screening do nothing to undermine the claim that treaties constrain state behavior as well. Empirically, we question von Stein's estimator on several grounds, including its strong distributional assumptions and its statistical inconsistency. We then illustrate that selection bias does not account for much of the difference between Simmons's (2000) and von Stein's (2005) estimated treaty effects, and instead reframe the problem as one of model dependency. Using a preprocessing matching step to reduce that dependency, we then illustrate treaty effects that are both substantively and statistically significant—and that are quite close in magnitude to those reported by Simmons.

Serious researchers in political science are finally beginning to take international legal agreements as worthy of sustained and rigorous analysis. Within the last several years, a growing group of scholars is making progress toward understanding the extent to which international law—and most specifically, the highly public and legal form of commitment represented in treaties—can actually shape the decisions governments make as well as broader outcomes of normative concern. The theory these studies draw on is becoming more refined: increasingly scholars are willing to analyze international legal agreements as a specific kind of commitment device. Treaties are the most formal “language” governments have to focus the expectations of individuals, firms, and other states that they seriously intend to keep their word in a particular policy area. Treaties enhance the reputational effects that may inhere in general policy declarations, precisely because they link performance to a broader principle that underlies the entire edifice of international law: *pacta sunt servanda*—treaties are to be observed. By choosing to become a treaty party, governments ante up a greater reputational stake than would otherwise be the case.

Estimating treaty effects is no simple thing, however. Despite terrific progress in supplementing case studies with quantitative models that test the generality of the claim that legal commitments matter, the evidentiary hurdles and methodological issues are highly contested. The most common worry is that treaty effects are merely reflections of underlying state preferences

rather than evidence of an independent influence on behavior (Downs, Rocke, and Barsoom 1996). This assertion is indeed troubling for those who would like to believe that governments can be nudged—if only at the margins—toward internationally preferred behaviors by making explicit agreements.

Jana von Stein<sup>1</sup> has made an important and sophisticated statistical contribution in this regard. Her strategy is to adapt a Heckman selection model to reestimate the impact of signing onto Article VIII, the section of the International Monetary Fund's (IMF) Articles of Agreement that prohibits signatories from restricting their current accounts. As a result, she argues that we should revise our estimates of the treaty's effect downward. Although Simmons (2000, 831) found that the marginal effect of signing onto Article VIII can be up to 27 percentage points in the second year after the last restriction, von Stein revises that estimate to 13 percentage points, although the estimated treaty effect remains both substantively and statistically significant.<sup>2</sup> Perhaps unintentionally, she also makes a contribution by showing that Simmons' original findings were sensitive to the strategy used to control for time, a point we develop here.

It would seem natural to apply a Heckman selection model to the potential problem of treaty selection bias. Our rejoinder is primarily cautionary. Choosing to attack selection bias<sup>3</sup> statistically rather than theoretically and empirically may account for selection “problems” without shedding much light on them. In statistical terms, von Stein argues that Article VIII is

Beth A. Simmons is Professor, Department of Government, Harvard University, 1033 Massachusetts Avenue, Cambridge, MA 02139 (bsimmons@latte.harvard.edu).

Daniel J. Hopkins is a Ph.D. student, Department of Government, Harvard University, Littauer Center, North Yard, Cambridge, MA 02138 (dhopkins@fas.harvard.edu).

The authors thank Adam Berinsky, Lynn Gale, Gary King, and Jasjeet Sekhon for statistical advice, three anonymous reviewers for valuable feedback, and Jana von Stein for helpful comments and assistance in replicating her model. All errors are our own. The data and code needed to replicate the analyses are available from the authors via email.

<sup>1</sup> All references to Jana von Stein refer to von Stein's article in this issue of *APSR*.

<sup>2</sup> As compared with Simmons (2000), von Stein also argues that the estimated treaty effects fade more quickly as the time from the most recent restriction passes. However, throughout this response, we emphasize the results in the first few years after the last restriction, as the data show these initial years to be the most critical in setting states on a restriction-free course.

<sup>3</sup> To be precise, by “selection bias,” we mean the bias resulting from the nonrandom assignment of the treatment stemming from both observable and unobservable sources.

not randomly assigned to countries even conditional on observed covariates, and that is indeed problematic. Theoretically, of course, this is to state the obvious. Random assignment would imply a theory of frivolous commitment-making, hardly a model on which a useful theory of compliance with legal obligations can be developed. We *know* treaty commitment is not random; that was shown in the original article (Simmons 2000). It does not follow that treaties are ineffectual. We view the process of making a treaty commitment as a costly policy that only a government with intentions to comply would generally be willing to make. *Ex ante*, for most governments, treaties involve ratification costs.<sup>4</sup> Government must have—or assemble—the basic political support to announce a change in legal regime for a particular policy. We should in most cases expect treaty ratification to be more costly *ex ante* than a mere policy announcement, because the ratification coalition will have to include not only those who may support the policy, but also those who want to tie the government's hands through altering the legal (and normative) setting in which policy is carried out. Because treaties focus expectations on compliance, ratifying a treaty without an intent to comply only raises *ex post* consistency costs. Indeed, the anticipation of such *ex post* costs should in fact contribute to the political opposition (hence, costs) a government faces *ex ante*. The bottom line is this: if treaties are commitment devices, then they *should* in fact have a screening effect, because only those governments that are willing and think they will be able to comply should sign on.

It is essential, however, to correct two implications of von Stein's discussion of treaty effects. First is the implication that anticipatory compliance casts doubt on the commitment value of the treaty itself. There is no reason to think this observed behavior undercuts a theory of the constraining power of treaties. Governments should rationally be concerned about the reputational costs of inconsistency. To move toward compliance prior to a formal commitment may reduce the uncertainty surrounding the ability to comply and is perfectly consistent with the theory advanced here. In fact, one shortcoming of the original article might have been to cast treaty effects too narrowly. If we include the anticipatory compliance treaties induce, we are likely to conclude Article VIII has an even more significant impact than Simmons (2000) originally reported.<sup>5</sup>

Second, and even more worrisome, von Stein's discussion suggests that screening effects and constraining effects are somehow mutually exclusive. We disagree.

<sup>4</sup> We are here using "ratification" in its broad political rather than narrow legal sense, although for some countries and issue areas they will be essentially the same.

<sup>5</sup> At the same time, we should also point out that for the subset of observations we employ in our following reestimation, these anticipated effects are far less pronounced than von Stein argues. The results presented in Table 2 are quite representative. When looking at the subset of signatories for which matched nonsignatories are available for the first of the matched datasets, the change in restriction behavior in the 4 years prior to signing is from restricting 70.5% of the time to restricting 65.9% of the time. The other matched datasets produce similar results.

Even for the committed, there may be conditions under which it would be tempting to renege on a treaty commitment. Many of these conditions will not have been fully anticipated by the government or indeed the ratifying coalition. But having paid the *ex ante* costs of ratification, a legally committed government will still rationally want to avoid the inconsistency costs of reneging. Our argument is that, *facing similar conditions*, Article VIII countries will try harder than will uncommitted countries to avoid restrictions, because they have staked their reputations on doing so. Screening and constraining are compatible treaty functions. The only real question is: how can we distinguish these effects empirically? As we show in the next section, the estimator offered by von Stein offers some advantages, but some serious drawbacks as well.

### HECKMAN SELECTION MODELS: A SOLUTION TO THE PROBLEM OF SELECTION BIAS?

Jana von Stein offers a potential solution to the problem of selection bias. She assumes that some of the important factors that explain selection into a treaty regime are unobservable and adapts a Heckman selection model to cope with that bias stemming from selection on unobservables. This section takes a close look at that choice. As is well known, Heckman models have some important limitations, and we demonstrate that, in this application, those limitations are pronounced. But we also believe von Stein has not conclusively isolated *selection effects*, and that much of the difference between her estimates and those in Simmons's (2000) original article are due to other specification choices. Having reframed the problem as one of model dependence, we go on to estimate the impact of Article VIII using techniques that markedly reduce model dependence—hence, that render more reliable results.

### Generic Issues

Heckman selection models<sup>6</sup> have enjoyed a recent burst of popularity in the political science literature (Berinsky 1999; Lemke and Reed 2001; Reed 2000; Timpone 1998; Vreeland 2003), although political methodologists are well aware of the problems with this class of models (Sartori 2003; Signorino 2003). Research has shown that Heckman-style models share several important weaknesses, including their sensitivity to specification, possible problems of collinearity, and heavy reliance on distributional assumptions (Lee 2001; Liao 1995; Sartori; Winship and Mare 1992). For precisely these reasons, recent methodological work on selection bias has focused on finding alternatives to the Heckman approach, often through semiparametric or nonparametric models (Heckman et al. 1998; Lee 2001; Sartori 2003; Winship and Morgan 1999). We

<sup>6</sup> For some of Heckman's initial work on selection bias, see Heckman 1976, 1979.

explore one such alternative, matching, in the final section of this article.

The problem of being overly reliant on distributional assumptions is a real issue here. In cases where the independent variables for the selection and outcome equations are the same, the standard Heckman selection model is identified *solely* on its distributional assumptions (Sartori 2003). To be sure, von Stein's model includes several variables that appear only in the selection equation, but no theoretical justification is given for why any of those variables is related to restriction behavior *only* through its impact on Article VIII commitment. In other words, it is unclear why any of those variables is a valid instrument with which to identify the model. We thus agree with Winship and Mare (1992, 342) who conclude that "Heckman's method is no panacea for selection problems and, when its assumptions are not met, may yield misleading results," a point that is also made by (Lee 1984). The problem of sensitivity to strong assumptions is especially pronounced in the case of von Stein's estimator, as she adds a second assumption of bivariate normality to a model that has already been criticized precisely for its dependence on distributional assumptions.

### An Inconsistent Estimator

An additional concern is specific to von Stein's adaptation of the Heckman probit. As she notes, she includes an indicator variable in the selection equation for observations that occur after a country has signed onto Article VIII. According to von Stein (??), the role of this indicator variable is to approximate survival analysis within a probit model by ensuring that the "estimated coefficients . . . are based only on the values of the independent variables before or in year  $t$ ," where " $t$ " refers to the year of signing. However, this indicator variable violates the non-quasi-complete separation assumption of logit and probit models: for probit models to render consistent estimates, they cannot include any independent variables that are perfect or quasi-perfect predictors of the dependent variable (Albert and Anderson 1984; Christmann and Rousseeuw 2001). Because observations only are coded as a "1" for this indicator variable if they have signed onto Article VIII—and are never coded as "1" when countries have not signed Article VIII—the indicator variable is a quasi-perfect predictor of the dependent variable. In such cases, there is no overlap between those observations that are predicted to be failures and those predicted to be successes; thus, the maximum likelihood estimates for the model's parameters do not exist. Some computer programs report parameter estimates under these conditions, but those estimates are not correct (Christmann and Rousseeuw 2001).

Put differently, the inclusion of this indicator variable means that even asymptotically, von Stein's estimator does not converge to the right estimates. One way to recognize this problem is to see if there are fitted values that only differ from 0 or 1 by tiny margins, and indeed, some 1,232 observations are predicted to sign with a probability above .999999. We confirmed

these suspicions using the Noverlap package in R 2.0.1 (R Development Core Team, 2004; Rousseeuw and Christmann 2004), which shows that there is no overlap when the indicator variable for signatories is included. The resulting inconsistency alone should constitute grounds to reject the estimated treaty effects von Stein presents.

### Explaining the Difference in Results: Selection Bias or Other Model Dependencies?

We have discussed the generic problems associated with Heckman selection models, and have argued that von Stein's adaptation produces estimates that are statistically inconsistent. Setting these issues aside, we now turn to whether von Stein has made a case that accounting for selection bias leads to a drastic revision of the estimated impact of Article VIII. Simmons (2000) estimated that Article VIII status makes the country on average 27 percentage points less likely to place restrictions, whereas von Stein estimates the treaty's actual *constraining* effect to be just 13 percentage points. von Stein (XX) attributes this gap *entirely* to selection bias: as she explains, "selection bias accounts for between 31% and 95% of the standard probit model's estimated effect of the legal commitment on a state's propensity to engage in compliant behavior." But von Stein has actually made several simultaneous changes to the original model, and only by disentangling them can we truly understand the extent to which Simmons's original estimate was driven by selection bias.

First, von Stein has changed the definition of the causal effect to be estimated. In a standard one-stage model, researchers often estimate causal effects by varying one or more independent variables while fixing the others to some value and then observing the difference in simulated values of the dependent variable under the model (King, Tomz, and Wittenberg 2000). In the model proposed by von Stein, we have separate estimates for the second-stage coefficients of the signatories and the nonsignatories. Instead of varying the values of key independent variables, then, von Stein fixes the values of all explanatory variables and varies the set of coefficients used to calculate the predicted probabilities. But if this is our strategy for estimating predicted probabilities, we need to specify *a priori* which values of the independent variables are of interest. Do we care about the effect of the treatment on the treated population, on the nontreated population, or on some other group?

von Stein chooses to focus on the effect of the treatment on the *nontreated*: she generates her estimates of the impact of Article VIII by measuring the change in the predicted probability that the mean nonsignatory will restrict its current account using first the nonsignatory and then the signatory outcome equations. But she could as easily have chosen to estimate the effect on the *treated* population instead. And in fact, when we estimate the treaty effect by focusing on its impact *on signatories*, we find that it is on average .04 larger for countries that restricted their current account in the

**TABLE 1. Estimated Treaty Effects**

Years Since Last Restriction	Selection Model, $\rho_s$ and $\rho_n = 0$		Selection Model, $\rho_s$ and $\rho_n$ Vary	
	Mean	95% Confidence Interval	Mean	95% Confidence Interval
0	.137	(.095,.183)	.098	(.059,.142)
1	.164	(.026,.308)	.129	(-.035,.285)
2	.023	(-.042,.093)	.012	(-.057,.088)
3	.021	(-.036,.081)	.012	(-.049,.077)
4	.020	(-.029,.072)	.011	(-.042,.067)
5	.018	(-.025,.065)	.011	(-.036,.059)
6	.017	(-.021,.059)	.010	(-.031,.054)
7	.016	(-.019,.055)	.010	(-.027,.049)

*Note:* von Stein makes several modifications to the model presented by Simmons (2000), and yet attributes the entire difference between her estimates and Simmons' estimates to selection bias. By fixing  $\rho_s$  and  $\rho_n$  at zero and then estimating the effect of signing Article VIII using the model presented by von Stein, we can observe how much impact selection bias—as opposed to other changes in how the effect is modeled—impact the estimates. The left columns present the estimated impact of Article VIII in a case where we have imposed the requirement of no selection bias, the right columns are our replication on von Stein's estimates. Making this direct comparison while holding other modeling decisions constant, we see that all else equal, selection bias has only a minor impact on our estimate of the treaty effect. Accounting for selection bias reduces the estimated effect from .137 to .098 in the first year since the last restriction, and from .164 to .129 in the second year.

past year than the results she reports.<sup>7</sup> There is nothing wrong with the choice to report on this relationship, but for comparative purposes, she is not reporting estimates on exactly the same causal relationship as that reported in Simmons 2000. For those who might wish to implement von Stein's model in the future, it is critical to specify *a priori* precisely the causal relationship in which they are most interested.

Another reason we cannot attribute the full difference in estimates to selection effects is that von Stein simultaneously switches from a logit to a probit functional form. This decision alone reduces the mean estimated Article VIII impact by .04 for countries that last restricted 1 year ago. Using the logit model, we estimate the marginal effect of signing Article VIII when a country is 1 year from its last restriction as .26, with a 95% confidence interval from .20 to .32. When switching to a probit model, however, the estimated mean marginal effect drops to .22, with a 95% confidence interval from .17 to .27. Again, we have no problem with this choice, and we recognize the probit is necessary to generate her specific selection model. Our point is simply that her results are driven to some extent by making different distributional assumptions, and not by the "problem" of selection bias.

The most substantial difference between von Stein's estimate and Simmons's (2000) estimate comes from how they deal with time.<sup>8</sup> If Simmons had accounted for time using two dummy variables in the same

way that von Stein does (for 0 and for 1 year since last restriction), the original article would have reported results that differ by only .007 from von Stein's for countries that had restricted in the prior year.<sup>9</sup> Simmons used splines instead (Beck, Katz, and Tucker 1998), a reasonable choice but not the only possible one. We are not accusing von Stein of handling the time dependence of observations inappropriately. But if Simmons had used two dummies in the original article, von Stein would not have had much of a case for a research note based on selection bias. Model dependencies, not selection bias, account for much of the gap between Simmons 2000 and von Stein.

Another way to show that selection bias may not account for the difference in results is by estimating separate probit models predicting restrictions for signatories and non-signatories. This is equivalent to estimating von Stein's selection model while fixing  $\rho_s$  and  $\rho_n$  both equal to zero. If von Stein is right that selection bias explains the majority of the change in the estimated effect, fixing  $\rho_s$  and  $\rho_n$  should lead the estimated Article VIII impact to return to something near its original estimate as presented in Simmons (2000). But as the similarity of the two estimates presented in Table 1 illustrates, that is far from the case. Imposing

for time rather than for splines. Beck, Katz, and Tucker (1998) note that there should not be any substantive difference on the estimate coefficients, and they mention that they have a slight preference for using splines over the dummies.

<sup>9</sup> Consider countries that are 1 year past their last restriction. The updated version of Simmons's model estimates the marginal effect of being an Article VIII signatory as .124, with 95% confidence intervals running from .08 to .17. The selection model predicts a highly similar marginal effect of .129, where the 95% confidence interval runs from -.04 to .29. The two estimates differ, clearly, in their uncertainty. But they provide nearly identical estimates of the mean Article VIII impact, a result that should cause us to be cautious in concluding that selection bias is what accounts for the differences between Simmons and von Stein.

<sup>7</sup> The marginal effect of the treaty on the mean nonsignatory is .10, with a 95% confidence interval from .06 to .14. For the mean signatory, though, the mean marginal effect increases to .14, with a 95% confidence interval from .08 to .19.

<sup>8</sup> As this note is primarily concerned about selection effects, we do not enter into an extended discussion about how correctly to handle the time series issues. The original 2000 article utilized a set of two cubic splines, generated using Beck, Katz, and Tucker's BTSCS program for STATA. von Stein chose to use two dummy variables to control

the condition of no selection bias, we observe estimated treaty impacts that are only slightly larger than those reported by von Stein. Hence other modeling decisions, including the switch from a logit to a two-stage probit model and to the use of dummy variables to control for time, must explain most of the reduction in the estimated treaty effect.

Even if we accept the smallest available estimates of the impact of Article VIII—those that come when Simmons's (2000) original model is modified to control for time using only two dummy variables instead of splines—two points are worth stressing. First, we dispute von Stein's conclusion that "a legal commitment to Article VIII appears to have little constraining power." Using the *smallest* estimates presented so far, 1 year after having restricted its current account, a country is likely to revert to restrictions with a probability of .24 if it is a nonsignatory, but just .11 if it has signed Article VIII. That is a change of over 50% in the probability of restricting, and it has proven quite robust—in not one of the specifications cited previously does the impact of the treaty become consistently statistically insignificant. Second, a convincing case has not been made that *selection bias* is what accounts for the bulk of the discrepancy between von Stein's results and those reported in Simmons. This becomes apparent only when we conduct a controlled methodological experiment and change one assumption at a time.

### CONTROLLING FOR BIAS: PROPENSITY SCORE MATCHING

Our theory of how and why international law works implies that governments do not enter into legal commitments randomly. If they did, commitments would hardly be credible and markets would have no reason to take Article VIII commitments seriously. The theory suggests that treaties screen *and* constraint. Von Stein's estimator does not convincingly show that the screening effects *overwhelm* the constraining effects of legal commitments. Nor does her statistical model advance our understanding of the factors that lead governments simultaneously to commit and to comply with their legal obligations, because the bias is attributed partly to "unobservables." But we do accept that the inherent problem of selection bias is potentially very real and must be addressed. Only by doing so will skeptics warm up to the idea that treaties not only screen but also constrain governments' future behavior.

We advocate the following. *Commitment* should be modeled by using the event history style of analysis employed in the original 2000 article. Every effort should be made to theorize and to include in the commitment model all observables theory suggests are relevant, and an effort should be made to theorize and measure purported "unobservables" as well. And to estimate the treaty's *effect* on subsequent behavior, we advocate matching techniques informed by both theory, and by the analysis of the decision to commit to the treaty. Nonparametric approaches such as matching control for bias on observables without making the strong distributional assumptions required by Heckman-type

models. And in recent work, they have demonstrated their utility when confronting thorny problems related to nonrandom assignment to treatment as well. (Harding 2003; Imai, 2005).

Our point of divergence from von Stein is our contention that important influences on commitment and compliance can be theorized, observed, and (imperfectly) measured. The most reasonable "unobservable" for which we agree it would be desirable to control is a government's *political will* to remove restrictions from the current account. If a government truly is determined to liberalize its economy, then we should be able to find traces of this in policy areas distinct from but related to the current account. We should expect a government that is intent on a program of liberalization—independent from its Article VIII commitment—to implement other policies designed to liberalize trade and to encourage the freer international movement of capital. A number of observable measures of political will can be used in this context. We use three. First, a government that has opened up its economy to capital flows likely has the "political will" to become integrated into the world economy. Second, a government that has become a member of the General Agreement on Tariffs and Trade (GATT, which evolved in 1995 into the World Trade Organization, or WTO) is also likely to have some "political will" to liberalize. And finally, a government that is more democratic might pursue economic openness and eschew restrictions that deny free access to foreign exchange.<sup>10</sup> Democracy was included in the original (Simmons 2000) model of commitment, and found not to be a strong influence. When we reran the original model, we found that countries that had opened their capital account were highly likely (hazard ratio = 5.33;  $p = .007$ ); GATT/WTO members were possibly likely (hazard ratio = 2.05;  $p = .24$ ); and democracies were less likely a positive influence on Article VIII adoption (hazard ratio = 1.05;  $p = .38$ ). A case can be made that these measures for political will should be taken into consideration when trying to determine the effect of Article VIII on the probability of restricting the current account.

In this section, we report the effects of Article VIII on restriction behavior estimated after a pre-processing matching step (Ho et al. 2004). Matching prior to performing standard parametric analyses reduces or eliminates the bias caused by selection on observable characteristics.<sup>11</sup> It also helps reduce the model dependency of our estimated effects, which is especially important given the sensitivity of the estimated effects to modeling decisions that we illustrated earlier. Using matching prior to implementing variants of the parametric model in Simmons 2000, we recover estimated treaty effects—defined as the average treatment effect—that are large and robust to model specification. For instance, the average treatment effect

<sup>10</sup> The measure used is the difference between democracy scores and autocracy scores, Polity IV dataset.

<sup>11</sup> For more on the theoretical foundations of matching, see Abadie and Imbens 2004, Imbens 2004, Rosenbaum and Rubin 1984.

for countries in the year after signing year is a reduction of 24.2 percentage points, with a 95% confidence interval from 3.9 percentage points to 43.1 percentage points. That is substantially closer to the estimated treaty effects of Simmons (2000) than those of von Stein. Simply put, Article VIII signatories are more likely to comply than are nonsignatories that are identical across a wide range of observed variables, including variables designed to proxy for political will.

To perform the matching, we first redefined our unit of observation to be a 6-year period of time during which we observe a country, or a "country-period." The 66 treated observations are countries that were Article VIII signatories for the first time in the fifth year of the observation window.<sup>12</sup> This allows us to observe the countries for 4 years prior to signing, for the signing year, and for 1 year following the signing year.<sup>13</sup> The universe of possible control cases includes all 6-year country-periods that do not overlap with the treated observations, for a total of 1,634 potential control cases. For instance, if Algeria never signs over the period of the dataset, but is observed for 30 years, it offers 25 possible control cases, one for each continuous 6-year period. And if Bangladesh signs in 1995, but was observed for 22 years before signing, the 13 prior periods that do not overlap with the treated observation period might provide potential control cases in those early years. This redefinition of the unit of observation would seem to markedly reduce the amount of data available to the researcher, but in fact it more accurately captures the transitions that we actually wish to observe—as well as the time-dependent structure of the observations.

Because this redefinition relies on combining 6 years of data, the data become far more sensitive to listwise deletion as a missing data strategy. We decided, then, to impute the missing covariates rather than discard the entire unit of observation (King et al. 2001) whenever data were missing. To do so, we used the *mix* package (Schafer 2003) in R 1.9.1. As a result, we have not one but five datasets, hence, five sets of matched observations. Estimating our causal effects across the five separate datasets allows us to incorporate the uncertainty that results from the imputation.

We then estimated a propensity score for each country-period in the new dataset using *MatchIt* (Ho et al. 2004b). A propensity score is the conditional probability of receiving the treatment (Rosenbaum and Rubin 1984)—that is, signing on to Article VIII after the fourth year—given the observed covariates. Doing so, we find that concerns about selection on *observables* are well justified: the mean propensity score

for the signatories is approximately .43, whereas for the control cases, it is just .02. Most of the cases in our prospective control group have a very low conditional probability of receiving the treatment. In other words, they are simply not comparable to the treated cases. They are highly unlikely to sign onto Article VIII during the observation window, and thus not very useful in estimating treatment effects.

To identify well-matched control cases from the 1,634 candidates, we then followed the guidelines proposed by Ho et al. (2004a). We matched exactly on the single most important predictor, the average number of years of restrictions placed in the 4-year pre-treatment observation period, and also matched on the estimated propensity score to achieve approximate balance on other covariates.<sup>14</sup> Of the 66 treated cases, we were able to match between 42 and 47 depending on the imputed sample. Initially, we tested for balance by ensuring that there were no significant differences in the treated and control samples on any of the 19 important covariates.<sup>15</sup> There were none. We looked for imbalanced samples by comparing all the possible multiplicative interactions of the 19 variables across the five matched datasets, and found just five significant differences out of the 1,805 possibilities. We then generated a list of other potentially unbalanced covariates by running sample *t*-tests and also the more powerful bootstrap Kolmogorov–Smirnov test<sup>16</sup> using the "Matching" package (Sekhon 2005) on all available Year 1 and Year 4 covariates. Any covariate whose *p*-value was under .10 on either test for any of the five matched datasets was noted. In all, the samples proved quite well balanced, with just 3 to 10 of the 37 covariates unbalanced for a given matched dataset. Not only do we have balance on most of the important predictors used by Simmons (2000) and von Stein, but we also have balance on our new measures of "political will" (capital account openness and GATT/WTO membership). Any estimated treatment effects, then, should be less vulnerable to concerns about political will as an omitted variable. And before running any parametric models, we have already identified those remaining confounders that could threaten our inferences.

<sup>14</sup> Even after matching, the treated and control groups differ in their propensity scores, so we used a caliper of .25 standard deviations to ensure that treated observations were not being matched to very different control-group observations.

<sup>15</sup> Here, "significantly different" is defined as any case where the *t* statistic comparing the means of the matched treatment and control groups is greater than 2. The "important covariates" are those that proved useful in estimating the propensity score. For Year 4 of the observation window—the year just prior to signing—these include restriction behavior, reserve volatility, reserves as a fraction of GDP, terms of trade volatility, GDP growth, GDP per capita, IMF surveillance, regional noncompliance, the use of IMF credits, the country's democracy score, its capital account openness, its status as a GATT/WTO member, and the calendar year. Also included in this test were the years of restriction-free behavior prior to the first year of the observation window, Year 1 gross national product (GNP) per capita, the number of years from joining the IMF to the beginning of the observation window, the use of IMF credits in Year 1, and the propensity score.

<sup>16</sup> For more on this test and its application to matched data, see Abadie 2002, Sekhon and Diamond 2005, and Sekhon 2004.

<sup>12</sup> The dataset covers countries through 1997 and includes 100 transitions into signatory status, but only 66 countries are observed across the 6-year observation window.

<sup>13</sup> Both von Stein's work and our own suggests that these 6 years are the most crucial window to isolate treaty effects. von Stein illustrates that beginning at roughly 4 years before signing, restriction behavior begins to change, and both von Stein and Simmons (2000) find that the strongest impact of Article VIII is in the first few years after last placing a restriction. Here, we measure the effects as they vary across a unit of time that is different and perhaps more intuitive. Rather than looking at the effects by when countries placed their last restriction, we look at the effects as the time from signing increases.



**TABLE 2. P-Values for *t*-Tests and Bootstrap Kolmogorov-Smirnov Tests**

	Mean, Treated	Mean, Control	<i>t</i> Test <i>p</i> -Value	KS Test <i>p</i> -Value
<b>Year 1 Covariates</b>				
Restrictions	0.705	0.727	0.816	
Years Since Last Restriction	2.159	2.455	0.793	0.976
Flexibility	1.386	1.455	0.523	
GNP/Capita	1976.245	1312.445	0.211	0.28
Change in GDP	3.663	3.925	0.896	0.604
Reserves/GDP	0.154	0.093	0.168	0.798
Reserve Volatility	-3.19	-3.209	0.906	0.758
Year	1987.023	1987.909	0.52	0.769
Terms of Trade Volatility	3.046	3.18	0.341	0.167
Universal	42.675	43.504	0.4	0.763
Regional Restrictions	34.574	27.698	0.275	0.062
IMF Surveillance	1.864	1.909	0.507	
Openness	78.148	82.117	0.758	0.058
GATT/WTO Member	1.705	1.705	1	
Balance of Payments/GDP	-4.353	-4.187	0.922	0.925
Use of Fund Credits	1.818	1.795	0.79	
Years of IMF Membership	25.159	24.795	0.887	0.498
<b>Year 4 Covariates</b>				
Restrictions	0.659	0.659	1	
Years Since Last Restriction	2.864	3.273	0.761	0.742
Flexibility	1.409	1.614	0.056	
GNP per Capita	2369.863	1958.404	0.491	0.916
Change in GDP	4.843	5.781	0.698	0.597
Reserves per GDP	0.172	0.159	0.781	0.301
Reserve Volatility	-3.186	-3.234	0.771	0.763
Terms of Trade Volatility	2.991	3.142	0.279	0.113
Universal	48.485	49.482	0.607	0.574
Regional Restrictions	42.14	36.61	0.361	0.092
IMF Surveillance	1.909	1.932	0.698	
Openness	79.596	81.073	0.908	0.308
GATT/WTO Member	1.75	1.773	0.805	
Balance of Payments/GDP	-4.444	-4.001	0.856	0.636
Use of Fund Credits	1.386	1.386	1	
Democracy	2.609	1.21	0.334	0.561
Capital Account Openness	1.134	1.114	0.772	0.945
<b>Other Covariates</b>				
Average Years of Restrictions, Years 1-4	0.705	0.705	1	1
Restrictions, Year 2	0.727	0.705	0.816	
Restrictions, Year 3	0.727	0.727	1	
Propensity Score	0.328	0.32	0.87	1

*Note:* This table presents *p*-values for *t*-tests and bootstrap Kolmogorov-Smirnov tests for the first of our five matched datasets. As is evident, the matched dataset is balanced on a wide range of covariates. Defining balance as all *p*-values higher than .05, these samples are balanced. Using our stricter criterion of *p* > .10, only four covariates are potentially unbalanced. And having identified those potential confounders, we can include them as explanatory variables in standard parametric models.

As an example of the balance obtained, consider the first of the five matches: across the 4 years prior to the potential treatment, the treated group and control group are identical in terms of their restriction behavior in Years 3 and 4, and differ only very slightly and insignificantly in Years 1 and 2. Or consider the calendar years of the matched groups. For the treatment group, the average year when the observation window begins is 1987, whereas for the control group it is 1988. As Table 2 makes clear, similarly close matches hold for the vast majority of variables in both Years 1 and 4 of the observation window. And of the 361 possible multiplicative interactions among the 19 key covariates, not

one has a *t* score greater than 2. For the matched countries that sign onto Article VIII within the observation window, we have identified control cases that are indistinguishable across a wide range of measures, from the presence or absence of IMF surveillance to the countries' GNP per capita. In terms of all of the variables we have observed in the pretreatment phase, the matched pairs differ chiefly in that the treated group actually signed on to Article VIII in the fifth year, whereas the control group did not sign on during the observation window.

What, then, is the estimated effect of signing Article VIII? Here, we used probit models to calculate the

average treatment effect for both the year of signing and the year after signing. Starting with Simmons's (2000) original model, we discarded those explanatory variables that were no longer approaching significance as predictors and tested those variables that were potentially unbalanced in any of the five matched samples as described earlier (in this case, Year 4 flexibility, Year 4 openness, and Year 1 regional restrictions). In the end, we wound up with a model of restriction behavior that closely approximated that presented by Simmons, with measures of openness in Year 1, democracy in Year 4, GATT/WTO status in Year 4, reserve volatility in Year 4, and an indicator variable marking whether the country last restricted in the previous year.<sup>17</sup> The estimated treaty effect in the signing year is a reduction of 17.7 percentage points, with a 95% confidence interval that runs from -0.7 percent points to 35.6 percentage points. For the year after signing, the same model leads to an estimated effect of 24.2 percentage points, with a 95% confidence interval from 3.9 percentage points to 43.1 percentage points. And as Ho et al. (2004) argue, using matching to preprocess the data reduces model dependency, and so should provide readers with added confidence that these results are not very sensitive to the specification of the parametric model, a point that our own data analyses confirm. These estimates are quite similar to those reported by Simmons, and confirm yet again that these estimated treaty effects are quite robust: they show up consistently across a wide range of modeling approaches and specifications.

To be sure, von Stein's critique is about nonrandom assignment to treatment owing to both observable and unobservable selection factors, and matching assumes that there is no selection on unobserved covariates. Certainly, though, matching can play a role in narrowing the range of possible unobservables, just as we demonstrated earlier. And it can also help in another way, even though it does not quantify the degree of selection on unobservables. By providing us with paired lists of countries that are highly similar on the observed covariates, matching allows us to draw upon knowledge not quantified within the data.<sup>18</sup> Instead of assuming that latent variables are distributed in a bivariate normal fashion—which is certainly not an assumption that researchers can observe in practice—we are making the more tangible assumption that each of our treated countries is similar in all important respects to its matched control. If there is indeed some unobservable influence, whether it is political will or anything else, careful study of the paired list in combination with substantive knowledge of the cases will help us understand what it might be. This is precisely what we advocate scholars do. We think this approach will yield far more insights into the selection effects in making international legal and other kinds of commitments than will fragile statistical methods that allow theoretically interesting processes to go unobserved.

<sup>17</sup> That is, we adopt the strategy for time dependence employed by von Stein, although we drop the second indicator variable, as it is zero across all observations.

<sup>18</sup> Ho et al. direct readers to (Rosenbaum 2002) chapter 3 for more on this point.

## REFERENCES

- Abadie, Alberto. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association* 97: 284–92.
- Abadie, Alberto, and Guido Imbens. 2004. "Large Sample Properties of Matching Estimators for Average Treatment Effects." [http://emlab.berkeley.edu/users/imbens/sme-21\\_jan\\_04.pdf](http://emlab.berkeley.edu/users/imbens/sme-21_jan_04.pdf) (July 12, 2005).
- Albert, A., and J. A. Anderson. 1984. "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika* 71 (1): 1–10.
- Beck, N., J. N. Katz, and R. Tucker. 1998. Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable. *American Journal of Political Science* 42 (4): 1260–88.
- Berinsky, Adam J. 1999. "The Two Faces of Public Opinion." *American Journal of Political Science* 43 (4): 1209–30.
- Christmann, Andreas, and Peter J. Rousseeuw. 2001. "Measuring Overlap in Logistic Regression." *Computational Statistics and Data Analysis* 37: 65–75.
- Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. "Is the good new about compliance good news about cooperation?" *International Organization* 50 (3): 379–406.
- Harding, David J. 2003. "Counterfactual Models of Neighborhood Effects: The Effect of Neighborhood Poverty on Dropping Out and Teenage Pregnancy." *American Journal of Sociology* 109 (3): 676–719.
- Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator of Such Models." *The Annals of Economic and Social Measurement* 5: 475–92.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.
- Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–98.
- Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2004a. "Matching as Nonparametric Preprocessing for Improving Parametric Causal Inference." <http://gking.harvard.edu/files/abs/matchup-abs.shtml> (January 12, 2005).
- Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2004b. "Matchit: Nonparametric Preprocessing for Parametric Causal Inference, Version 1.0-1." Harvard University, Cambridge, MA. Computer program available at <http://gking.harvard.edu/matchit/>.
- Imai, Kosuke. 2005. Do Get-Out-The-Vote Calls Reduce Turnout?: The Importance of Statistical Methods for Field Experiments. *American Political Science Review* 99 (2): 283–300.
- Imai, Kosuke, Gary King, and Olivia Lau. 2005. "Zelig: Everyone's Statistical Software, Version 2.2-2." Harvard University, Cambridge, MA. Computer program available at <http://gking.harvard.edu/zelig/>.
- Imbens, Guido. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity. A Review." *Review of Economics and Statistics* 86 (1): 4–30.
- King, Gary, James Honaker, Anne Joseph, and Kenneth F. Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95 (1): 49–69.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 41–55.
- Lee, Lung-fei. 1984. "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity." *Econometrica* 52 (4): 843–63.
- Lee, Lung-fei. 2001. "Self-Selection." In *A Companion to Theoretical Econometrics*, ed. B. H. Baltagi. Malden, MA: Blackwell Publishers Inc.
- Lemke, Douglas, and William Reed. 2001. "War and Rivalry Among Great Powers." *American Journal of Political Science* 45 (2): 457–69.
- Liao, Tim Futing. 1995. "The Nonrandom Selection of Don't Knows in Binary and Ordinal Responses: Corrections with the Bivariate Probit Model with Sample Selection." *Quality and Quantity* 29: 87–110.

- R Development Core Team. 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44 (1): 84–93.
- Rosenbaum, Paul E., and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79 (387): 516–24.
- Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York, NY: Springer Verlag.
- Rousseeuw, Peter J., and Andreas Christmann. 2004. "The Noverlap Package," version 1.0–1. Available online at <http://cran.r-project.org/doc/packages/noverlap.pdf>.
- Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions." *Political Analysis* 11: 111–38.
- Schafer, Joseph. 2003. "The Mix Package: Estimation/Multiple Imputation for Mixed Categorical and Continuous Data." <http://cran.r-project.org/src/contrib/Descriptions/mix.html> (January 14, 2005).
- Sekhon, Jasjeet S. 2004. "The Varying Role of Voter Information across Democratic Societies." <http://jskhon.fas.harvard.edu/papers/SekhonInformation.pdf> (February 24, 2005).
- Sekhon, Jasjeet S. 2005. "Matching: Multivariate and Propensity Score Matching with Automated Balance Search." Computer program available at <http://jskhon.fas.harvard.edu/matching>.
- Sekhon, Jasjeet, and Alexis Diamond. "Genetic Matching for Estimating Causal Effects: A New Method of Achieving Balance in Observational Studies." Presented at the Summer Meeting of the Society for Political Methodology, Tallahassee, FL.
- Signorino, Curt. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11 (4): 316–44.
- Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94 (4): 819–35.
- Timpone, Richard J. 1998. "Structure, Behavior, and Voter Turnout in the United States." *American Political Science Review* 92 (1): 145–58.
- Vreeland, James Raymond. 2003. *The IMF and Economic Development*. New York: Cambridge University Press.
- Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18: 327–50.
- Winship, Christopher, and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25: 659–707.

## APSR EXTERNAL REVIEWERS, 2004–2005

*The peer review process relies on the professionalism and generosity of those who contribute their time to read and evaluate the work of others. The Editor thanks the following scholars, who served as manuscript reviewers between August 15, 2004 and August 14, 2005.*

Arash Abizadeh  
 Alan I. Abramowitz  
 Paul R. Abramson  
 Martha Ackelsberg  
 James Adams  
 Peter J. Ahrens Dorf  
 Scott Ainsworth  
 John H. Aldrich  
 John Alford  
 James E. Alt  
 Scott L. Althaus  
 R. Michael Alvarez  
 Edwin Amenta  
 Christopher J. Anderson  
 Elizabeth S. Anderson  
 Stephen D. Ansolabehere  
 Enriqueta Aragonés  
 Kevin Arceneaux  
 Asher Arian  
 Larry Arnhart  
 Theodore S. Arrington  
 Scott Ashworth  
 David Austen-Smith  
 Deborah Avant

Amy Baehr  
 Michael A. Bailey  
 Lisa Baldez  
 Richard J. Ball  
 Steven J. Balla  
 Susan A. Banducci  
 Navin A. Bapat  
 Katherine Barbieri  
 Joel D. Barkan  
 Michael Barnett  
 David P. Baron  
 Charles J. Barrilleaux  
 Clyde W. Barrow  
 Jens Bartelson  
 Gad Barzilai  
 James S. C. Battista  
 Lawrence Baum  
 Matthew A. Baum  
 Frank R. Baumgartner  
 Deborah Baumgold  
 Kathleen Bawn  
 Brady P. Baybeck  
 Clive Bean  
 David H. Bearce  
 Nathaniel Beck  
 Caroline C. Beer  
 Ronald Beiner  
 Aaron Belkin  
 Eva Bellin  
 Jonathan Bendor

Sara C. Benesh  
 Andrew Bennett  
 D. Scott Bennett  
 Kenneth Benoit  
 Richard F. Bensel  
 Michelle A. Benson  
 Pablo Beramendi  
 Jeffrey Berejikian  
 Adam Berinsky  
 Michael Berkman  
 William T. Bernhardt  
 William D. Berry  
 William T. Bianco  
 Susan Bickford  
 Jack Bielasiak  
 Glen Biglaiser  
 Sarah A. Binder  
 Sarah Birch  
 Colin Bird  
 Benjamin G. Bishin  
 Harry Blair  
 André Blais  
 Shannon Lindsey Blanton  
 Francis Bloch  
 Frederick J. Boehmke  
 Iris Bohnet  
 John Bohte  
 Kenneth Bollen  
 Wilko Bolt  
 Gary E. Bolton  
 Jon R. Bond  
 Richard Born  
 Stephen A. Borrelli  
 Shaun Bowler  
 Janet Box-Steffensmeier  
 Paul R. Brace  
 Ted Brader  
 Henry E. Brady  
 Steven J. Brams  
 Patrick T. Brandt  
 Paul Brass  
 Michael Bratton  
 John Brehm  
 Peter Breiner  
 Paul Brewer  
 Stephen G. Brooks  
 Bryan Brophy-Baermann  
 Robert D. Brown  
 Trevor Brown  
 J. Lawrence Broz  
 Allen Buchanan  
 Bruce Bueno de Mesquita  
 Ethan Bueno de Mesquita  
 Barry C. Burden  
 Katrina Burgess

Ross E. Burkhart  
 Nancy Burns  
 Paul Burstein  
 Marc L. Busch  
 Tim Büthe  
 Melissa A. Butler  
 Daniel L. Byman

Luis Cabrera  
 Bruce E. Cain  
 Thomas M. Callaghy  
 Steven J. Callander  
 Randall L. Calvert  
 Charles M. Cameron  
 Maxwell A. Cameron  
 James E. Campbell  
 Damaris Canache  
 Brandice Canes-Wrone  
 Bradley C. Canon  
 David T. Canon  
 Nicholas Capaldi  
 John M. Carey  
 Daniel P. Carpenter  
 Clifford James Carrubba  
 Thomas M. Carsey  
 Jamie L. Carson  
 Carol A. Cassel  
 James W. Ceaser  
 Lars-Erik Cederman  
 Simone Chambers  
 Eric C. C. Chang  
 Henry W. Chappell, Jr.  
 Evan Charney  
 Kerry A. Chase  
 Jeffrey Checkel  
 Pradeep Chhibber  
 Thomas Christiansen  
 Michael Suk-Young Chwe  
 Claudio Cioffi-Revilla  
 Jack Citrin  
 Harold D. Clarke  
 Cornell W. Clayton  
 Elisabeth S. Clemens  
 Joshua Clinton  
 Robert L. Clinton  
 Michael D. Cobb  
 David Cochran  
 Molly Cochran  
 Susan D. Collins  
 Patricia D. Conley  
 John Conybeare  
 Martin L. Cook  
 Alexander Cooley  
 Laurence D. Cooper  
 Brian Copeland  
 Dale Copeland

Michael J. Coppedge  
 Javier Corrales  
 Anne N. Costain  
 Gary W. Cox  
 Stephen C. Craig  
 Neta C. Crawford  
 Markus Crepaz  
 Ingrid Creppell  
 Mark J. C. Crescenzi  
 Ann N. Crigler  
 Brian F. Crisp  
 Christophe M. Crombez  
 Audrey Kurth Cronin

Xinyuan Dai  
 Ernesto Dal Bo  
 Fred R. Dallmayr  
 Russell J. Dalton  
 Vesna Danilovic  
 Christian Davenport  
 Christina Davis  
 Darren Davis  
 Sue Davis  
 Suzanna De Boef  
 Scott de Marchi  
 Christopher J. Deering  
 Thomas S. Deluca, Jr.  
 Karl DeRouen  
 Uday Desai  
 Louis DeSipio  
 Scott W. Desposato  
 David Dessler  
 Alberto Diaz-Cayeros  
 Bruce Dickson  
 Paul F. Diehl  
 Joshua Foa Dienstag  
 Henry Dietz  
 Mary G. Dietz  
 Avinash Dixit  
 William J. Dixon  
 David Dollar  
 Richard Doner  
 Han Dorussen  
 Keith L. Dougherty  
 Keith Dowding  
 William M. Downs  
 Michael W. Doyle  
 James N. Druckman  
 John S. Dryzek  
 Raymond M. Duch  
 John S. Duffield  
 John Duggan  
 Thomas L. Dumm  
 Thad Dunning  
 Raymond D. Duvall

Kent Eaton  
 George C. Edwards, III  
 Sean Ehrlich  
 Eldon Eisenach  
 Todd Eisenstadt  
 Peter Eisinger  
 Stephen L. Elkin

Tanja Ellingsen  
 Colin Elman  
 Peter Emberley  
 Walter Enders  
 Stephen G. Engelmann  
 Erik J. Engstrom  
 Andrew J. Enterline  
 Robert M. Entman  
 David Epstein  
 Lee Epstein  
 Hülya K. Eraslan  
 Robert S. Erikson  
 Maria Escobar-Lemmon  
 Kevin M. Esterling  
 Margarita Estevez-Abe  
 J. Peter Euben  
 C. Lawrence Evans  
 Diana Evans  
 Geoffrey Evans

Tulia G. Falletti  
 James Farr  
 Cynthia Farrar  
 David M. Farrell  
 James D. Fearon  
 Richard E. Feinberg  
 Scott L. Feld  
 Stanley Feldman  
 Sven Feldmann  
 John A. Ferejohn  
 Kathy Ferguson  
 Jan Fidrmuc  
 Ada W. Finifter  
 Steven E. Finkel  
 Martha Finnemore  
 Markus Fischer  
 Steven P. Forde  
 Benjamin O. Fordham  
 Richard Fording  
 Jonathan A. Fox  
 Fabio Franchino  
 Ronald A. Francisco  
 Paul Franco  
 Charles H. Franklin  
 Robert J. Franzese, Jr.  
 Guillaume R. Fréchette  
 Paul Freedman  
 John R. Freeman  
 Kim L. Fridkin  
 Jeffry A. Frieden  
 Elisabeth Friedman  
 H. Richard Friman  
 Paul Froese  
 Norman Frohlich  
 Timothy M. Frye  
 Paul Frymer  
 Dieter Fuchs  
 Timothy Fuller  
 Carolyn L. Funk

Matthew Gabel  
 Brian Gaines  
 Vincenzo Galasso

William A. Galston  
 Robert T. Gannett, Jr.  
 James C. Garand  
 Geoffrey M. Garrett  
 Bryan Garsten  
 Scott Sigmund Gartner  
 Mark J. Gasiorowski  
 Scott Gates  
 Kurt Taylor Gaubatz  
 Gerald F. Gaus  
 Claudine Gay  
 Barbara Geddes  
 John G. Geer  
 Andrew Gelman  
 Christopher F. Gelpi  
 Dino Gerardi  
 Alan Gerber  
 Elisabeth Gerber  
 Deborah J. Gerner  
 John Gerring  
 Hazem Ghobarah  
 James L. Gibson  
 Martin Gilens  
 Micheal W. Giles  
 Anthony J. Gill  
 Howard Gillman  
 John Gilmour  
 James G. Gimpel  
 James M. Glaser  
 Garrett Glasgow  
 James M. Glass  
 Amihai Glazer  
 Kristian S. Gleditsch  
 Dennis W. Gleiber  
 Gary D. Glenn  
 Gary Goertz  
 Sona Nadenichek Golder  
 Joshua S. Goldstein  
 Kenneth M. Goldstein  
 Ewa A. Golebiowska  
 Michael E. Goodhart  
 Robert E. Goodin  
 Sanford C. Gordon  
 William T. Gormley, Jr.  
 Peter A. Gourevitch  
 Joanne Gowa  
 Mark A. Graber  
 Robert C. Grady, II  
 Robert Grafstein  
 Ruth W. Grant  
 Donald Green  
 Kevin B. Grier  
 Bernard N. Grofman  
 Paul W. Gronke  
 Kimberly A. Gross  
 John Gruhl  
 Thomas Gschwend  
 John G. Gunnell  
 Dipak K. Gupta

Jacob S. Hacker  
 Frances Hagopian  
 Donald P. Haider-Markel  
 Susan B. Haire  
 Zoltan L. Hajnal

- |                       |                         |                       |
|-----------------------|-------------------------|-----------------------|
| Henry E. Hale         | John D. Huber           | Sadafumi Kawato       |
| Melinda Gann Hall     | Leonie Huddy            | Mark Andreas Kayser   |
| Peter A. Hall         | Valerie Hudson          | William R. Keech      |
| Richard L. Hall       | Simon Hug               | Philip Keefer         |
| Mark Hallerberg       | Kevin W. Hula           | Daniel Kelemen        |
| Daniel Hallin         | Wendy Hunter            | Judith Kelley         |
| James T. Hamilton     | Patricia A. Hurley      | Patrick J. Kenney     |
| Thomas H. Hammond     | Jon Hurwitz             | Marc Kilgour          |
| Jain Hampsher-Monk    | Mark Hurwitz            | Claire Jean Kim       |
| Ange-Marie Hancock    | Vincent Hutchings       | David C. Kimball      |
| Roger B. Handberg     | Paul K. Huth            | Desmond King          |
| Bruce E. Hansen       |                         | Gary King             |
| Susan B. Hansen       | Guido Imbens            | George Klosko         |
| Wendy L. Hansen       | Indridi Indridason      | Jack Knight           |
| Thomas Hansford       | Robert P. Inman         | Joseph M. Knippenberg |
| John W. Hardin        | Michael D. Intriligator | Jack H. Knott         |
| Robert Harmel         | Douglas Irwin           | Jeffrey W. Koch       |
| Bard Harstad          | Nancy Isenberg          | Stephen A. Kocs       |
| Jonathan Hartlyn      | John T. Ishiyama        | Gregory Koger         |
| Virginia Haufler      | Torben Iversen          | Ariel Kohen           |
| Emily Hauptmann       | Shanto Iyengar          | Ken W. Kollman        |
| Daniel Hausman        |                         | Barbara Koremenos     |
| Mary E. Hawkesworth   | Robert W. Jackman       | Walter Korpi          |
| Darren Hawkins        | Simon Jackman           | Thad Kousser          |
| Jude C. Hays          | Lawrence R. Jacobs      | Mathew Krain          |
| Håvard Hegre          | Gary C. Jacobson        | Isaac Kramnick        |
| Thomas Heilke         | William G. Jacoby       | George A. Krause      |
| Susan Hekman          | Alison M. Jaggar        | Keith Krehbiel        |
| Gretchen Helmke       | Patrick James           | Amie D. Kreppel       |
| Jeffrey R. Henig      | Richard Jankowski       | Glen S. Krutz         |
| Paul R. Hensel        | Ted G. Jelen            | Chandran Kukathas     |
| Jeffrey Herbst        | Jeffery A. Jenkins      | James H. Kuklinski    |
| Rodney E. Hero        | Erin Jenne              | Sheri Kunovich        |
| Richard K. Herrmann   | M. Kent Jennings        | Charles Kurzman       |
| Michael C. Herron     | Nathan Jensen           | Andrew Kydd           |
| Roberta Q. Herzberg   | Jennifer Jerit          | Will Kymlicka         |
| Gregory Hess          | Robert Jervis           |                       |
| Marc J. Hetherington  | Calvin C. Jillson       | Dean Lacy             |
| Virginia A. Hettinger | James D. Johnson        | Krishna K. Ladha      |
| John R. Hibbing       | Martin Johnson          | David D. Laitin       |
| Alexander M. Hicks    | Simon Johnson           | David A. Lake         |
| Benjamin Highton      | Timothy R. Johnson      | Gina Lambright        |
| Sunshine Hillygus     | Richard Johnston        | Dimitri A. Landa      |
| Nancy J. Hirschmann   | Bradford S. Jones       | Todd Landman          |
| Jack Hirshleifer      | Bryan D. Jones          | Pierre F. Landry      |
| Michael J. Hiscox     | David R. Jones          | Laura I. Langbein     |
| Simon Hix             | Mark P. Jones           | Laura Langer          |
| Matthew Hoddie        | Peter Jones             | David J. Lanoue       |
| Anke E. Hoeffler      | Michael A. Jones-Correa | Ted Lascher           |
| Valerie J. Hoekstra   | Henrik Jordahl          | Richard R. Lau        |
| Thomas J. Hoffman     | Will R. Jordan          | Michael Laver         |
| Marie E. Hojnacki     | Peter Josephson         | Jennifer L. Lawless   |
| Thomas Holbrook       | Jane Junn               | Eric D. Lawrence      |
| Charles A. Holt       |                         | Chappell Lawson       |
| Bonnie Honig          | Ronald Kahn             | Jeffrey R. Lax        |
| Christopher Hood      | Stathis N. Kalyvas      | Geoffrey C. Layman    |
| Ted Hopf              | Cynthia S. Kaplan       | David Leblang         |
| Alan Houston          | Noah Kaplan             | James H. Lebovic      |
| Christopher Howard    | Devesh Kapur            | Frances Lee           |
| Robert M. Howard      | Jonathan N. Katz        | Taeku Lee             |
| William Howell        | Richard S. Katz         | Beth L. Leech         |
| Evelyne Huber         | Stuart Kaufman          | Brett Ashley Leeds    |
| Gregory A. Huber      | Karen M. Kaufmann       | Douglas W. Lemke      |
|                       |                         | Gabriel Lenz          |
|                       |                         | Ross Levine           |

Jack S. Levy  
 Jacob T. Levy  
 Jeffrey B. Lewis  
 Michael S. Lewis-Beck  
 Quan Li  
 Mark I. Lichbach  
 Evan Lieberman  
 Michael Lienesch  
 Fernando Limongi  
 Tse-min Lin  
 Staffan Lindberg  
 Stefanie Lindquist  
 Christian List  
 Sharon Anne Lloyd  
 Milton G. Lodge  
 John Londregan  
 Andrew G. Long  
 Burdett A. Loomis  
 David L. Lowery  
 Robert C. Lowry  
 David Lublin  
 David Ludden  
 Pauline Jones Luong  
 Arthur W. Lupia  
 Robert C. Luskin  
 Ian S. Lustick  
 Ellen Lutz  
 Cecelia Lynch

Paul K. MacDonald  
 Stephen Macedo  
 Gerald L. Mackie  
 Michael B. MacKuen  
 Susan A. MacManus  
 Stephen J. Majeski  
 Forrest Maltzman  
 Jane Mansbridge  
 Edward D. Mansfield  
 Harvey C. Mansfield, Jr.  
 Zeev Maoz  
 Gerald M. Mara  
 George E. Marcus  
 Michael Margolis  
 Patchen Markell  
 Gary Marks  
 Jonathan Marks  
 Monty Marshall  
 Andrew D. Martin  
 Cathie Jo Martin  
 Lanny Martin  
 Paul S. Martin  
 Wendy Martinek  
 Michael D. Martinez  
 T. David Mason  
 John Matsusaka  
 James A. McCann  
 Nolan M. McCarty  
 Harwood McClerking  
 Kirstie McClure  
 Scott D. McClurg  
 Mathew D. McCubbins  
 Monika L. McDermott  
 Eileen Lorenzi McDonagh

Michael P. McDonald  
 Michael McFaul  
 Anthony J. McGann  
 Fiona McGillivray  
 Kathleen McGraw  
 Kevin T. McGuire  
 John P. McIver  
 Timothy J. McKeown  
 Lauren M. McLaren  
 Iain S. McLean  
 Michael McLendon  
 Andrew McLennan  
 Kathleen R. McNamara  
 Walter R. Mebane, Jr.  
 John Norman Medearis  
 Michael F. Meffert  
 Bonnie Meguid  
 Kenneth J. Meier  
 Adam Meirowitz  
 Arthur M. Melzer  
 Samuel Merrill, III  
 Carol A. Mershon  
 Lee Kendall Metcalf  
 Joel S. Migdal  
 Sidney M. Milkis  
 Arthur H. Miller  
 Helen Milner  
 William Mishler  
 Joshua M. Mitchell  
 Sara McLaughlin Mitchell  
 Timothy P. Mitchell  
 Subrata Mitra  
 Devra Coren Moehler  
 Karl Ove Moene  
 John Mollenkopf  
 Jeffery J. Mondak  
 Sarah Monoson  
 Burt L. Monroe  
 Nathan Monroe  
 Jose G. Montalvo  
 Alexander H. Montgomery  
 Gabriella R. Montinola  
 J. Donald Moon  
 Will H. Moore  
 Massimo Morelli  
 Scott Morgenstern  
 James A. Morone  
 James D. Morrow  
 Rebecca B. Morton  
 Robert G. Moser  
 Michael A. Mosher  
 Layna Mosley  
 Shaheen Mozaffar  
 Gary Mucciaroni  
 Anthony Mughan  
 J. Russell Muirhead  
 Vincent Phillip Muñoz  
 James C. Murdoch  
 Shoon Murray  
 Diana C. Mutz  
 David Myatt  
 Roger B. Myerson  
 Jack H. Nagel

Jonathan Nagler  
 Peter F. Nardulli  
 Patrick Neal  
 Cary J. Nederman  
 Thomas E. Nelson  
 W. Russell Neuman  
 Eric Neumayer  
 Richard G. Niemi  
 Tim Nokken  
 Timothy Nokken  
 Tim Nordstrom  
 Helmut Norpoth  
 Barbara Norrander  
 William Nylen

Thomas H. Oatley  
 John S. Odell  
 Sharyn O'Halloran  
 Rosemary H. T. O'Kane  
 J. Eric Oliver  
 John R. Oneal  
 Barry O'Neill  
 John Orbell  
 Clifford Orwin  
 Laurence J. O'Toole, Jr.  
 Zoe M. Oxley  
 Caglar Ozden

Alexander C. Pacek  
 Robert Pahre  
 Glenn Palmer  
 Michael J. Palmer  
 Thomas L. Pangle  
 Philip Paolino  
 Roland Paris  
 Craig Parsons  
 Alan W. Patten  
 Mark Peceny  
 Randall Peerenboom  
 Mark A. Peffley  
 Elizabeth Maggie Penn  
 Anibal Perez-Linan  
 M. J. Peterson  
 Mark A. Peterson  
 Jon C. Pevehouse  
 Robert J. Phillips  
 Daniel Philpott  
 Patrick A. Pierce  
 Jennifer Pitts  
 Eric Plutzer  
 Steven C. Poe  
 Solomon W. Polachek  
 Mark A. Pollack  
 Brian M. Pollins  
 Jonas Pontusson  
 Daniel Posner  
 Marsha Pripstein Posusney  
 Matthew Potoski  
 G. Bingham Powell, Jr.  
 Robert Powell  
 Timothy J. Power  
 Andrea Prat  
 Richard Price  
 Robert Price



- Vincent Price  
David M. Primo  
Markus Prior  
Lant Pritchett  
Elisabeth M. Prugl  
Adam Przeworski  
Ethan Putterman
- Dennis P. Quinn  
Kevin Quinn  
Paul J. Quirk
- George Rabinowitz  
Benjamin Radcliff  
Lyn Ragsdale  
J. Mark Ramseyer  
Robert Rauchhaus  
James Lee Ray  
David P. Redlawsk  
Patrick M. Regan  
Andrew Rehfeld  
Mark Reinhardt  
William M. Reisinger  
Dan Reiter  
Thomas F. Remington  
Nicholas Rengger  
Rafael Reuveny  
Tom W. Rice  
Michael J. Rich  
Bradley M. Richardson  
Shelley Rigger  
Jonathan M. Riley  
Evan J. Ringquist  
Jeffrey M. Ritter  
Brian E. Roberts  
Kenneth Roberts  
David B. Robertson  
James A. Robinson  
Jonathan Rodden  
Philip G. Roeder  
James R. Rogers  
Ronald Rogowski  
Robert Rohrschneider  
Susan Rose-Ackerman  
Richard Rosecrance  
Frances M. Rosenbluth  
Peter Rosendorff  
Howard Rosenthal  
Robert S. Ross  
Lawrence S. Rothenberg  
Peter L. Rousseau  
Dane Rowlands  
Thomas J. Rudolph  
Nita Rudra  
Jerrold G. Rusk  
Bruce M. Russett
- Andrew Sabl  
Scott D. Sagan  
Sebastian M. Saiegh  
Brian Sala  
Nicholas Sambanis  
Lynn M. Sanders
- M. Elizabeth Sanders  
Todd Sandler  
Arlene W. Saxonhouse  
Denise Schaeffer  
Andreas Schedler  
Ethan Scheiner  
Kim Lane Scheppele  
William Scheuerman  
Kenneth Scheve  
Eric Schickler  
Wendy Schiller  
Kay Lehman Schlozman  
Holger Schmidt  
Stephen Schneck  
Ben Ross Schneider  
Gerald Schneider  
Mark Schneider  
John T. Scholz  
Christian Schultz  
Kenneth A. Schultz  
Melissa Ann Schwartzberg  
Jillian Schwedler  
Randall Schweller  
James C. Scott  
John T. Scott  
Lyle Scruggs  
John Seaman  
Amy Searight  
Jeffrey A. Segal  
Gary M. Segura  
AnnaMaria Seleny  
Mitchell A. Seligson  
Holli Semetko  
Itai Sened  
Paul Senese  
Patrick Seyd  
Dhavan Shah  
Michal Shamir  
Ian Shapiro  
Robert Y. Shapiro  
Ira Sharkansky  
Elaine B. Sharp  
Daron R. Shaw  
Colleen Sheehan  
Stephen Shellman  
Kenneth Shepsle  
Kenneth S. Sherrill  
Charles Shipan  
W. Phillips Shively  
Kenneth W. Shotts  
Matthew Soberg Shugart  
John Sides  
Kathryn Sikkink  
Joel Silbey  
Beth A. Simmons  
Randy Simmons  
Adam Simon  
Barbara Sinclair  
Valeria Sinclair-Chapman  
David A. Singer  
Randolph M. Siverson  
Theda Skocpol  
Branislav L. Slantchev
- Oleg Smirnov  
Alastair Smith  
Eric R. A. N. Smith  
Jay M. Smith  
Jeffrey Smith  
Kevin B. Smith  
Mark A. Smith  
Rogers M. Smith  
Steven B. Smith  
Steven S. Smith  
Regina A. Smyth  
Duncan Snidal  
David A. Snow  
Jack L. Snyder  
James M. Snyder, Jr.  
Donald R. Songer  
Arndt Sorge  
David Soskice  
Joe Soss  
Bartholomew H. Sparrow  
Pablo T. Spiller  
Jeff Spinner-Halev  
James F. Spriggs, II  
Hendrik Spruyt  
Allan C. Stam, III  
Marco R. Steenbergen  
Robert E. Stein  
Peter J. Steinberger  
Sven Steinmo  
John D. Stephens  
Randolph T. Stevenson  
Charles Stewart, III  
Marianne C. Stewart  
James A. Stimson  
Shannon C. Stimson  
Susan Stokes  
Clarence N. Stone  
Walter J. Stone  
Jeffrey M. Stonecash  
Randall W. Strahan  
Håvard Strand  
Dara Z. Strolovitch  
Kaare Strøm  
Terry Sullivan  
Duane H. Swank  
Todd Swanstrom
- Charles S. Taber  
Charles Tarlton  
G. Alan Tarr  
Sidney G. Tarrow  
Andrew J. Taylor  
Scott Taylor  
Michelle M. Taylor-Robinson  
Benno Teschke  
Paul E. Teske  
Strom C. Thacker  
Elizabeth Theiss-Morse  
Sean M. Theriault  
Leslie Paul Thiele  
Raju G. C. Thomas  
Alexander Thompson  
William R. Thompson  
Charles Tilly  
Michael M. Ting

Jaroslav Tir  
James Tong  
Roger Tourangeau  
Zev Trachtenberg  
John Transue  
Daniel S. Treisman  
James L. True  
Roy T. Tsao  
Joshua A. Tucker  
Alex Tuckness  
Mark Tunick  
Mark V. Tushnet  
Tom Tyler  
  
Benjamin A. Valentino  
Nicholas A. Valentino  
Nicolas Van de Walle  
Jan W. Van Deth  
Stephen W. Van Evera  
Robert P. van Houweling  
Georg S. Vanberg  
Tatu Vanhanen  
Miguel Vatter  
Lynn Vavreck  
Richard Vengroff  
Lisa Pace Vetter  
Erik Voeten  
David Vogel  
Craig Volden  
Jana von Stein  
James Raymond Vreeland

R. Harrison Wagner  
Paul J. Wahlbeck  
Kenneth D. Wald  
Jeremy Waldron  
John R. Wallach  
Michael Wallerstein  
Katherine Cramer Walsh  
Barbara F. Walter  
Vincent Wei-cheng Wang  
Leonard Wantchekon  
Paul Wapner  
Michael D. Ward  
Mark E. Warren  
Paul V. Warwick  
Stephen L. Wasby  
Richard Waterman  
William L. Waugh  
Gregory J. Wawro  
Christopher R. Way  
Georgina Waylen  
Frank Whelon Wayman  
M. Stephen Weatherford  
Leonard B. Weinberg  
Margaret Weir  
Herbert F. Weisberg  
Susan Welch  
Alan Wertheimer  
Kurt Weyland  
Stephen D. Whitefield  
Paul F. Whiteley  
Guy Whitten

Keith E. Whittington  
Erik M. Wibbels  
Clyde Wilcox  
Kenneth C. Williams  
Melissa Williams  
Rick K. Wilson  
Donald Wittman  
Christopher Wlezien  
William C. Wohlforth  
Harold Wolman  
Brantly Womack  
Meredith Woo-Cumings  
B. Dan Wood  
John T. Woolley  
Anne Wren  
Gerald C. Wright  
John R. Wright  
  
Bernard Yack  
David Yalof  
Garry Young  
Iris Marion Young  
  
John R. Zaller  
David H. Zarefsky  
Langche Zeng  
Lawrence Ziring  
Alan S. Zuckerman  
Michael P. Zuckert

## INDEX TO VOLUME 99

- Abizadeh, Arash. Does Collective Identity Presuppose an Other? On the Alleged Incoherence of Global Solidarity... 45
- Alexandrova, Anna, see Slantchev, Branislav L. . . . . 459
- Alford, John R., Carolyn L. Funk, and John R. Hibbing. Are Political Orientations Genetically Transmitted? . . . . . 153
- Bahry, Donna, Mikhail Kosolapov, Polina Kozyreva, and Rick K. Wilson. Ethnicity and Trust: Evidence from Russia . . . . . 521
- Bailey, Michael A. Welfare and the Multifaceted Decision to Move . . . . . 125
- Basinger, Scott J., and Howard Lavine. Ambivalence, Information, and Electoral Choice . . . . . 169
- Baybeck, Brady, see Berry, William D. . . . . 505
- Berry, William D., and Brady Baybeck. Using Geographic Information Systems to Study Interstate Competition . . . . . 505
- Bianco, William T., and Itai Sened. Uncovering Evidence of Conditional Party Government: Reassessing Majority Party Influence in Congress and State Legislatures . . . . . 361
- Doyle, Michael W. Three Pillars of the Liberal Peace . . . . . 463
- Falleti, Tulia G. A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective . . . . . 327
- Frymer, Paul. Racism Revised: Courts, Labor Law, and the Institutional Construction of Racial Animus . . . . . 373
- Funk, Carolyn L., see Alford, John R. . . . . 153
- Gartzke, Erik, see Slantchev, Branislav L. . . . . 459
- Gerber, Alan S., and Donald P. Green. Correction to Gerber and Green (2000), Replication of Disputed Findings, and Reply to Imai (2005) . . . . . 301
- Gerring, John, Strom C. Thacker, and Carola Moreno. Centripetal Democratic Governance: A Theory and Global Inquiry . . . . . 567
- Goeree, Jacob K., and Charles A. Holt. An Explanation of Anomalous Behavior in Models of Political Participation . . . . . 201
- Gordon, Sanford C., and Catherine Hafer. Flexing Muscle: Corporate Political Expenditures as Signals to the Bureaucracy . . . . . 245
- Grant, Ruth W., and Robert O. Keohane. Accountability and Abuses of Power in World Politics . . . . . 29
- Green, Donald P., see Gerber, Alan S. . . . . 301
- Gunnell, John G. Political Science on the Cusp. Recovering a Discipline's Past . . . . . 597
- Hafer, Catherine, see Gordon, Sanford C. . . . . 245
- Hibbing, John R., see Alford, John R. . . . . 153
- Hill, Kim Quaile, and Tetsuya Matsubayashi. Civic Engagement and Mass-Elite Policy Agenda: Agreement in American Communities . . . . . 215
- Holt, Charles A., see Goeree, Jacob K. . . . . 201
- Hopkins, Daniel J., see Simmons, Beth A. . . . . 623
- Imai, Kosuke. Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments . . . . . 283
- Jacobs, Lawrence R., and Benjamin I. Page. Who Influences U.S. Foreign Policy? . . . . . 107
- Kayser, Mark Andreas. Who Surfs, Who Manipulates? The Determinants of Opportunistic Election Timing and Electorally Motivated Economic Intervention . . . . . 17
- Kedar, Orit. When Moderate Voters Prefer Extreme Parties: Policy Balancing in Parliamentary Elections . . . . . 185
- Keohane, Robert O., see Grant, Ruth W. . . . . 29
- King, Desmond S., and Rogers M. Smith. Racial Orders in American Political Development . . . . . 75
- Kinsella, David. No Rest for the Democratic Peace . . . . . 453
- Koremenos, Barbara. Contracting around International Uncertainty . . . . . 549
- Kosolapov, Mikhail, see Bahry, Donna . . . . . 521
- Kozyreva, Polina, see Bahry, Donna . . . . . 521
- Laver, Michael. Policy and the Dynamics of Political Competition . . . . . 263
- Lavine, Howard, see Basinger, Scott J. . . . . 169
- Leach, William D., and Paul A. Sabatier. To Trust an Adversary: Integrating Rational and Psychological Models in Collaborative Policymaking . . . . . 491
- Leventoglu, Bahar, and Ahmer Tarar. Prenegotiation Public Commitment in Domestic and International Bargaining . . . . . 419
- Lieberman, Evan S. Nested Analysis as a Mixed-Method Strategy for Comparative Research . . . . . 435
- Martin, Lanny W., and Georg Vanberg. Coalition Policymaking and Legislative Review . . . . . 93
- Matsubayashi, Tetsuya, see Hill, Kim Quaile . . . . . 215
- Meguid, Bonnie M. Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success . . . . . 347
- Mitzen, Jennifer. Reading Habermas in Anarchy: Multilateral Diplomacy and Global Public Spheres . . . . . 401
- Moreno, Carola, see Gerring, John . . . . . 567
- Mutz, Diana C., and Byron Reeves. The New Videomalaise: Effects of Televised Incivility on Political Trust . . . . . 1
- Page, Benjamin I., see Jacobs, Lawrence R. . . . . 107
- Putterman, Ethan. Rousseau on the People as Legislative Gatekeepers, Not Framers . . . . . 145
- Reeves, Byron, see Mutz, Diana C. . . . . 1
- Robertson, David Brian. Madison's Opponents and Constitutional Design . . . . . 225
- Rosato, Sebastian. Explaining the Democratic Peace? . . . . . 467
- Rosendorff, B. Peter. Stability and Rigidity: Politics and Design of the WTO's Dispute Settlement Procedure . . . . . 389
- Rueda, David. Insider-Outsider Politics in Industrialized Democracies: The Challenge to Social Democratic Parties . . . . . 61
- Sabatier, Paul A., see Leach, William D. . . . . 491
- Scott, John T. Rousseau's Anti-Agenda-Setting Agenda and Contemporary Democratic Theory . . . . . 137
- Sened, Itai, see Bianco, William T. . . . . 361
- Simmons, Beth A. and Daniel J. Hopkins. The Constraining Power of International Treaties: Theory and Methods . . . . . 623
- Slantchev, Branislav L., Anna Alexandrova, and Erik Gartzke. Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace . . . . . 459
- Slantchev, Branislav L. Military Coercion in Interstate Crises . . . . . 533
- Smith, Rogers M., see King, Desmond S. . . . . 75
- Song, Sarah. Majority Norms, Multiculturalism, and Gender Equality . . . . . 473
- Stokes, Susan C. Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina . . . . . 315
- Tarar, Ahmer, see Leventoglu Bahar . . . . . 419
- Thacker, Strom C., see Gerring, John . . . . . 567
- Vanberg, Georg, see Martin, Lanny W. . . . . 93
- von Stein, Jana. Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance . . . . . 611
- Whittington, Keith E. "Interpose Your Friendly Hand": Political Supports for the Exercise of Judicial Review by the United States Supreme Court . . . . . 583
- Wilson, Rick K., see Bahry, Donna . . . . . 521

# **journals in Political Science from Cambridge**

## **British Journal of Political Science**

British Journal of Political Science is a broadly based journal aiming to cover developments across a wide range of countries and specialisms. Contributions are drawn from all fields of political science (including political theory, political behaviour, public policy and international relations), and articles from scholars in related disciplines (sociology, social psychology, economics and philosophy) appear frequently. With a reputation established over 30 years of publication, British Journal of Political Science is widely recognized as one of the premier journals in its field.

quarterly, volume 36, 2006, ISSN 007-1234

Institutions, print + online: \$330

Institutions, online-only: \$295

Individuals, print only: \$49

**Editors:**

**SARAH BIRCH,**

**DAVID SANDERS,**

**HUGH WARD, and ALBERT WEALE,**

**University of Essex, UK**

## **Studies in American Political Development**

Studies in American Political Development publishes scholarship on political change and institutional development in the United States from a variety of theoretical viewpoints. Articles focus on governmental institutions over time and on their social, economic and cultural setting. In-depth presentation in a longer format allows contributors to elaborate on the complex patterns of state-society relations. SAPD encourages an interdisciplinary approach and recognizes the value of comparative perspectives.

**Editors:**

**KAREN ORREN,**

**University of California, Los Angeles**

**STEPHEN SKOWRONEK,**

**Yale University**

semi-annual, volume 20, 2006, ISSN 0898-588x

Institutions, print + online: \$158

Institutions, online-only: \$133

Individuals, print only: \$56

**CAMBRIDGE  
UNIVERSITY PRESS**

40 west 20th street  
new york, ny 10011-4211

tel: 800.872.7423

fax: 845.353.4141

**journals\_subscriptions@cup.org • www.journals.cambridge.org**